

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900
Perfect score: 101
Sequence: 1 cgtgcacatccgcgcgcatgg.....gcgtgcgcgcgcacacatc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST: *
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.4	33.1	434	10	BE122146
2	33	32.7	632	10	BE582676
3	32.6	32.3	921	9	AL574395
4	32.2	31.9	1114	14	BQ050843
5	32	31.7	476	9	AA024701
6	31.8	31.5	421	9	AV163491

c	7	31.8	31.5	448	10	AV644360
c	8	31.8	31.5	452	9	AU094715
c	9	31.8	31.5	495	10	AV640754
c	10	31.8	31.5	497	10	AV628507
c	11	31.8	31.5	503	10	AV628086
c	12	31.8	31.5	507	10	AV638132
c	13	31.8	31.5	511	10	AV631846
c	14	31.8	31.5	517	10	AV391885
c	15	31.8	31.5	517	10	AV635168
c	16	31.8	31.5	534	10	AV621642
c	17	31.8	31.5	536	10	AV396567
c	18	31.8	31.5	635	13	B1874098
c	19	31.8	31.5	656	13	B1724994
c	20	31.8	31.5	666	14	BQ810877
c	21	31.8	31.5	677	12	BG860320
c	22	31.6	31.3	1300	14	BM919318
c	23	31.4	31.1	311	9	AA853014
c	24	31.4	31.1	427	9	AA668638
c	25	31.4	31.1	431	12	BF111987
c	26	31.4	31.1	451	12	BF111254
c	27	31.4	31.1	474	12	BG492131
c	28	31.4	31.1	512	9	AA516134
c	29	31.4	31.1	535	9	AA522923
c	30	31.4	31.1	585	12	BF435325
c	31	31.4	31.1	1630	10	BE421606
c	32	30.6	30.3	462	9	AL822466
c	33	30.6	30.3	499	10	BE497906
c	34	30.6	30.3	718	14	BQ841046
c	35	30.6	30.3	847	9	BQ841544
c	36	30.6	30.3	847	9	AL569731
c	37	30.4	30.1	272	12	BG057161
c	38	30.4	30.1	336	13	BM662318
c	39	30.4	30.1	427	12	BG223323
c	40	30.4	30.1	456	9	AA705536
c	41	30.4	30.1	513	10	AW168364
c	42	30.4	30.1	629	14	BM971132
c	43	30.4	30.1	629	14	BQ183984
c	44	30.2	29.9	260	10	AV640218
c	45	30.2	29.9	435	13	BM259173

ALIGNMENTS

RESULT 1
BE122146/c 434 bp mRNA linear EST 13-JUN-2000
LOCUS
DEFINITION 894017F07.Y2 C. reinhardtii CC-1690, normalized, lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BE122146
VERSION
KEYWORDS
SOURCE
ORGANISM
EST

REFERENCE 1 (bases 1 to 434)
Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
McDermott, J. P., Silflow, C., Stern, D. and Surzycki, R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 2

JOURNAL
COMMENT
Contact: Elizabeth H. Harris
Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu

FEATURES
source
1..434
/organism="Chlamydomonas reinhardtii"

```
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:1055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda zap
II"
/Note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, contains cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light. TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
POLYA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into Lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
ZAP clones by superinfection with ExSist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      67 a      120 c      163 g      84 t
ORIGIN
Query Match      33.1%; Score 33.4; DB 10; Length 434;
Best Local Similarity 60.4%; Pred. No. 28;
Matches 55; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 5 GCAATCGCCGATGATAGGCGCCGAGTGTTCGCCACGCGCCCTCAAGTGGC 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 364 GCCCAGCGGACCTGTGATGGGAGCGG;GTTCACCACCTTGCCACGACGAGC 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 65 AGCGTTACAGCCATGGGCGCGCGCGC 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 304 ATCAGTGTCTGTCCATGCGCCACGACGCGCAGC 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
BE582676      632 bp      mRNA      linear      EST 16-NOV-2000
LOCUS      BE582676
DEFINITION      9-10D-WY Projeamr Phytophthora sojae cDNA, mRNA sequence.
ACCESSION      BE582676
VERSION      BE582676.1 GI:9833624
KEYWORDS      EST.
SOURCE      Phytophthora sojae.
ORGANISM      Eukaryota; stramenopiles; Comycetes; Pythiales; Pythiaceae;
      Phytophthora.
REFERENCE      1 (bases 1 to 632)
      Qutob,D., Hraber,P.T., Sobral,B.W.S. and Gijzen,M.
      Comparative analysis of expressed sequences in Phytophthora sojae
      Plant Physiol. 123 (1), 243-254 (2000)
JOURNAL      20267956
COMMENT      Contact: Gijzen M
      Agriculture and Agri-Food Canada
      1391 Sandford Street, London, Ontario, Canada N5V 4T3
      Tel: 519 457 1470
      Fax: 519 457 3997
      Email: gijzenm@agr.ca
      Location/Qualifiers:
        1..632
        /organism="Phytophthora sojae"
        /strain="race 2, strain P6497"
        /db_xref="taxon:67493"
        /clone_lib="PsojaeIV"
        /dev_stage="mycelium"
        /lab_host="E. coli strain XL0R"
        /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI. This
        cDNA library was constructed from POLY(A+) enriched mRNA
        from mycelium grown in liquid medium for 3 weeks at 25 C
        in the dark in a liquid synthetic medium of 2.4 g sucrose
        , 0.27 g asparagine, 10 mg cholesterol, 10 mg ascorbate,
        and 2 mg thiamine, per liter, plus salts and minerals.
        Complementary DNA was synthesized from mRNA using an
        xhoI-poly(dT) linker-primer. EcoRI adapters were ligated
        to the blunt-ended cDNA fragments and the products were
        digested with xhoI for directional cloning into lambda ZAP
```

```
Express vector. This lambda library was amplified once
using E. coli host strain XL1 Blue MRF+. Inserts were then
subcloned by mass excision using ExSist helper phage for
conversion into phagemid vector pBK-CMV in E. coli host
strain XL0R. Sequenced using T3 primer: 5' ATT AAC CCT
CAC TAA AGG GA 3'."
BASE COUNT      157 a      172 c      192 g      105 t      6 others
ORIGIN
Query Match      32.7%; Score 33; DB 10; Length 632;
Best Local Similarity 59.6%; Pred. No. 38;
Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 4 TGCATATCGCCGATGATAGGCGCCGAGGTGTTCGCCACGCGCCCTCAAGTGG 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9 TCCAGATCGCCAGGCGCGTGGCGCCGAGGTGATCCGACGTCACGACGACGACGTGG 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 64 CAGCGTTACAGCCATAGGCGC 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 69 AGCTGTCAAGTGTGCGGCG 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AL574395      921 bp      mRNA      linear      EST 16-FEB-2001
LOCUS      AL574395
DEFINITION      AL574395 LRI_NFL006.PL2 Homo sapiens cDNA clone CS0D1058XP20 3
      prime, mRNA sequence.
ACCESSION      AL574395
VERSION      AL574395.1 GI:12934563
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 921)
      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
      Full-length cDNA libraries and normalization
      JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
      Genoscope - Centre National de Sequencage
      BP 191 91006 Evry cedex - France
      Email: segref@genoscope.cns.fr; Web : www.genoscope.cns.fr.
      Location/Qualifiers:
        1..921
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="CS0D1058XP20"
        /clone_lib="LRI_NFL006.PL2"
        /tissue_type="Placenta"
        /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
        was primed with a NotI-oligo(dT) primer. Five prime end
        enriched, double-stranded cDNA was digested with Not I and
        cloned into the Not I and Eco RV sites of the pCMVSPORT 6
        vector. Library was normalized. Library was constructed by
        life technologies. Contact : Feng Liang Life Technologies,
        a division of Invitrogen 9800 Medical Center Drive
        Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
        Email: fliang@lifestech.com URL:
        http://fulllength.invitrogen.com."
BASE COUNT      151 a      265 c      344 g      145 t      12 others
ORIGIN
Query Match      32.3%; Score 32.6; DB 9; Length 921;
Best Local Similarity 59.6%; Pred. No. 52;
Matches 53; Conservative 1; Mismatches 35; Indels 0; Gaps 0;

OY 13 CCCGATGATAGGCGCCGAGGTGTTCGCCACGCGCCGCTCAAGTGGCGCGTTC 72
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 87 CCCCGTGGCGGCGGCGGAGGTGCGAGCGCCAGCGCGTGTCCAGTGAAGCGGTGC 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 73 AGCCATGAGCGTGGCGCGCAGCAGCATC 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 147 TGGCGTGTGGAGCCCGGCGCGCTTAGC 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


[illegible]

TITLE	B., Morris,M., Parsons,J.T., Prange,C., Rikkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tran,F., Thierly-Meg,J., Treviski,E., Underwood,K., Wohlmann,P., Waterson,R., Wilson,R. and Matra,M.
JOURNAL	Generation and analysis of 280,000 human expressed sequence tags
MEDLINE	Genome Res. 6 (9), 807-828 (1996)
COMMENT	97044478 Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watsn.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 593 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 365. Location/Qualifiers 1..476 /organism="Homo sapiens" /db_xref="GDB:1281894" /db_xref="taxon:9606" /clone="IMAGE:365190" /clone_lib="Soares_fetal_heart_NDH19W" /sex="unknown" /dev_stage="19 weeks" /lab_host="DH10B (ampicillin resistant)" /note="Organ: heart; Vector: pT7/3D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCACTGTGAAGTGAGCGCCGCCATCTTTTATTTTTTTTT 3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7/3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5 library constructed by M.Patrina Bonaldo. This library was constructed from the same fetus as the fetal lung library. Soares fetal lung NBdh19W."
BASE COUNT	74 a 126 c 191 g 79 t 6 others
ORIGIN	
Query Match	31.7%; Score 32; DB 9; Length 476;
Best Local Similarity	59.6%; Pred. No. 68;
Matches	53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY	13 CCCGATGGATGAGGGCCGAGGTTCCGCACAGCGGAGCCGTCACAGTGGGCGAGCGGTTC 72
Dd	157 CCCCgggAGNAGAGGTGAGtgAGgtcAGGccCAGGGcgctcGtCgAgTAgtAGAcGgGgtGC 216
OY	73 AGGCCATGGCGGTGCCGCGCACGCACATC 101
Dd	217 TGGCGGCTGTGGAGCCCCGCGGTTCAC 245
RESULT 6	
LOCUS	AUI63491/c 421 bp mRNA linear EST 03-APR-2002
DEFINITION	AUI63491 Rice cDNA from young root Oryza sativa (japonica cultivar-group) cDNA clone R10853, mRNA sequence.
ACCESSION	AUI63491
VERSION	AUI63491
KEYWORDS	AUI63491.1 GI:11118303
SOURCE	EST.
ORGANISM	Oryza sativa (japonica cultivar-group).
	Oryza sativa (japonica cultivar-group).
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriharoidaeae; Oryzaceae; Oryza.
REFERENCE	1 (bases 1 to 421) Sasaki,T. and Yamamoto,K.
AUTHORS	Rice cDNA from young root (2000)
TITLE	Unpublished (2000)
JOURNAL	Contact: Takuji Sasaki
COMMENT	

National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8607, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsusaka@agr.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>
PROJECT = "NGP" (R15744444)

FEATURES

```

source
1. .421
/organism="Oryza sativa (japonica cultivar-group)"
/db_xref="taxdb:33947"
/clone_id="R10853"
/clone_lib="Rice cDNA from young root"
/tissue_type="young root"
69 a 136 c 138 f 73 t 5 others
BASE COUNT
ORIGIN

```

Query Match	31.5%	Score	31.8	DB	9	Length	421
Best Local Similarity	64.0%	Pred. No.	75				
Matches	48	Conservative	0	Mismatches	27	Indels	0
						Gaps	0

OY	20	GATGAGGGCCAGGTGTTCGCACGGCGAAGCCTCCAAAGTGGGACAGCTTCAGGCCAT	79
Dd	383	GATGAGGGCCCCGGTGCTCGTCCCAGGGAACACTCGAAGTAGTCGCGCATCTCGGCTC	324
OY	80	GGGCGTGCGCGCAC	94
Dd	323	CGGCGCGTCAAGCTC	309

RESULT 7

LOCUS	448 bp	mRNA	linear	EST 15-DEC-2000
AV644360				
DEFINITION	AV644360	Chlamydomonas reinhardtii 5' CO2	Chlamydomonas reinhardtii	
ACCESSION	AV644360	U00878.5		
		5' mRNA sequence.		

KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii

REFERENCE
1 (bases 1 to 448)
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

AUTHORS Asakimura, Y., Miura, K., Kuchic, K., Inoue, Y., Fukuzawa, H., Ohyama, K.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of *Chlamydomonas reinhardtii*
JOURNAL DNA Res. 7 (5), 303-307 (2010)

COMMENT

FEATURES

The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: [http://www.kazusa.or.jp/en/plant/
location/Qualifiers/](http://www.kazusa.or.jp/en/plant/location/Qualifiers/)

nos
son

```

/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db.xref="taxon:3055"
/clone="HCL087a05.1"
/clone_1lb="Chlamydomonas reinhardtii 5% CO2"
/notes="Vector: pLivescript1r SK-, Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"
BASE COUNT      75 a      124 c      161 g      88 t
BRGIN

```

Query Match	31.5%;	Score 31.8;	DB 10;	Length 448;
Best Local Similarity	59.3%;	Pred. NO. 76;		
Matches	54;	Conservative	0;	Mismatches 37;
			Indels	0;
			Gaps	0;

QY 5 GCAATGCCCCGAGGATAGGGGGCCGAGGTTTTCGCCACGGCAGCCGTCGCAAGTGGGC 64
Db 247 GCCACACGGGACCTTGATGGGGGAGCCGGTTCACCACTTGTGGCAGCAGCACCAAGACC 188
QY 65 AGCGTTTCAGGCCATGGCGTGGCGCCGCGCAGG 95
Db 187 ATCACTGCTGTCCATGGCCAGCGCACGCCAGG 157

RESULT 8

LOCUS	AUD094715	452 bp	mRNA	linear	EST 03-APR-2002
DEFINITION	AUD094715 Rice panicle (longer than 10cm) Oryza sativa (japonica cultivar-group) cDNA clone E20707, mRNA sequence.				
ACCESSION	AUD094715				
VERSION	AUD094715.1	GI:8857397			
KEYWORDS	EST.				

ORGANISM
Oryza sativa (Japanese cultivar-group)
Eukaryota: Viridiplantae: Streptophyta
Spermatophyta: Magnoliophyta: Liliopsida: Poales
Eumarioidae: Oryzae: Oryza.
1 (bases 1 to 452)

REFERENCE

NO.	TITLE	JOURNAL	COMMENT
1	Rice cDNA from panicle (longer than 10cm) (2000)	Unpublished (2000)	Contact: Takuji Sasaki National Institute of Advanced Industrial Science and Technology (AIST)

Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasakileabr.affrc.go.jp, URL:http://rjp.dna.affrc.go.jp/
PROJECT = 'RGP',
Country: Japan

nos

BASE COUNT	ORIGIN
78 a	150 c 139 g 85 t

Query Match	31.5%	Score	31.8	DB	9	Length	452
Best Local	Similarity	64.0%	Pred. No.	76			
Matches	48	Conservative	0	Mismatches	27	Indels	0
						Gaps	0

OY 20 GATAGGGGCGCAGGTTCGCCACGGCGATCCACTGGGACGGCTTCAAGGCAT 79
 ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 GATCAGCGCGCCGGTGCTCGTCCGGCGATTACACTGGAAGTAGTCGGACCTCGGCTC 361L

QY	80	GGGCGTGCCGCGCAC	94
Db	360	CGGCCCGTCCAGCTC	346

RESULT 9
AV6A0754

LOCUS	AV640754	495 bp	mRNA	linear	EST 15-DEC-2000
DEFINITION	AV640754 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii				
ACCESSION	cdna clone HCL021h03_5' , mRNA sequence.				
VERSION	AV640754				
	AV640754.1	GI:10784082			

SOURCE ORGANISM	CHLAMYDOMONAS
Chlamydomonas reinhardtii.	Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii	Chlamydomonas reinhardtii
Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;	Chlamydomonas
Chlamydomonadaceae; Chlamydomonas.	

REFERENCE 1 (bases 1 to 495)
AUTHORS Asamizu,E., Mura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
source 1..495
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL021h03_r"
/clone_lib="Chlamydomonas reinhardtii 5% CO2"
/note="Vector: pBluescriptII SK-; Site.1: EcoRI; Site.2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"
BASE COUNT 77 a 145 c 175 g 98 t
ORIGIN
Query Match 31.5%; Score 31.8; DB 10; Length 495;
Best Local Similarity 59.3%; Pred. No. 77;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 5 GCAAATCGCCCGATGATGATGAGGCGGCGAGGTTCGCCACGCGAGCCCGTCCAAATGGGCG 64
DB 367 GCCACCGCGACCTTGATGGGGAGCGGTGTTCACACCTTCTGCGACGACACGAGCC 308
QY 65 AGCGTTTCAGGCCATGGCGGTGCGCGCGACG 95
DB 307 ATCAGTGTCTGTCATGCGCACGACGCGACG 277
RESULT 10
AV628507/c 497 bp mRNA linear EST 15-DEC-2000
LOCUS AV628507 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL042h12_r 5', mRNA sequence.
ACCESSION AV628507 GI:10791141
VERSION AV628507.1 GI:10791141
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 497)
AUTHORS Asamizu,E., Mura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
source 1..497
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL042h12_r"
/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site.1: EcoRI; Site.2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon

dioxide concentration in the bubbling gas was changed from
5% to 0.04%"
BASE COUNT 77 a 148 c 174 g 98 t
ORIGIN
Query Match 31.5%; Score 31.8; DB 10; Length 497;
Best Local Similarity 59.3%; Pred. No. 77;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 5 GCAAATCGCCCGATGATGATGAGGCGGCGAGGTTCGCCACGCGAGCCCGTCCAAATGGGCG 64
DB 370 GCCACCGCGACCTTGATGGGGAGCGGTGTTCACACCTTCTGCGACGACACGAGCC 311
QY 65 AGCGTTTCAGGCCATGGCGGTGCGCGCGACG 95
DB 310 ATCAGTGTCTGTCATGCGCACGACGCGACG 280
RESULT 11
AV628086/c 498 bp mRNA linear EST 15-DEC-2000
LOCUS AV628086 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LCL036e01_r 5', mRNA sequence.
ACCESSION AV628086
VERSION AV628086
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 498)
AUTHORS Asamizu,E., Mura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
source 1..498
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="LCL036e01_r"
/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site.1: EcoRI; Site.2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"
BASE COUNT 78 a 148 c 174 g 98 t
ORIGIN
Query Match 31.5%; Score 31.8; DB 10; Length 498;
Best Local Similarity 59.3%; Pred. No. 77;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 5 GCAAATCGCCCGATGATGATGAGGCGGCGAGGTTCGCCACGCGAGCCCGTCCAAATGGGCG 64
DB 370 GCCACCGCGACCTTGATGGGGAGCGGTGTTCACACCTTCTGCGACGACACGAGCC 311
QY 65 AGCGTTTCAGGCCATGGCGGTGCGCGCGACG 95
DB 310 ATCAGTGTCTGTCATGCGCACGACGCGACG 280
RESULT 12
AV638132/c 503 bp mRNA linear EST 15-DEC-2000
LOCUS AV638132

DEFINITION	AV6318132 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii cDNA cloneHC082403_r 5', mRNA sequence.			
ACCESSION	AV6318132			
VERSION	AV6318132.1			
KEYWORDS	GI:10781452			
SOURCE	EST.			
ORGANISM	Chlamydomonas reinhardtii.			
REFERENCE	Chlamydomonas reinhardtii. Eukaryote: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.			
AUTHORS	1 (bases 1 to 503) Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Rabata,S.			
TITLE	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii			
JOURNAL	DNA Res. 7 (3), 305-307 (2000)			
MEDLINE	20539644			
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers 1..503 /organism="Chlamydomonas reinhardtii" /strain="C9" /db_xref="taxon:3355" /clone="HC082403_r" /clone_1p="Chlamydomonas reinhardtii 5% CO2" /note="Vector: pluescript SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"			
BASE COUNT	79 a 147 c 178 g 99 t			
ORIGIN				
Query Match	31.5%; Score: 31.8; DB 10; Length 503;			
Best Local Similarity	59.3%; Pred: No. 77;			
Matches	54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;			
OY	5 GCAAATGCCCGCATGGATGAGGGCCGAGACTGTTCGCCACGCGAGCCGCTCCAGTGGCC 64			
DB	367 GCCACCGGACACCTTGATGGGGGAGCCCGCTGTTCCACACCTTCTGGCCACGACGACGACG 308			
OY	65 AGCGGTTCAAGCGCATGGCGCGTGGCCGCG 95			
DB	307 ATCACTGCTGTCCATGGCCACGACGACGCG 277			
RESULT 13				
AV631846/c				
LOCUS	AV631846 507 bp mRNA linear EST 15-DEC-2000			
DEFINITION	AV631846 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone ICL10004_r 5', mRNA sequence.			
ACCESSION	AV631846			
KEYWORD	AV631846.1 GI:10794480			
ORGANISM	EST.			
SOURCE	Chlamydomonas reinhardtii.			
REFERENCE	Chlamydomonas reinhardtii. Eukaryote: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.			
AUTHORS	1 (bases 1 to 507) Asamizu,E., Miura,K., Kuchio,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Rabata,S.			
TITLE	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii			
JOURNAL	DNA Res. 7 (5), 305-307 (2000)			
MEDLINE	20539644			
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers			
FEATURES				

```

source
1. 507
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="fcl100g04.r"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI. The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"
BASE COUNT      83 a      149 c      176 g      99 t
ORIGIN
Query Match      31.5%; Score 31.8; DB 10; Length 507;
Best Local Similarity 59.3%; Pred. No. 77;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
5 GCAAATCGCCCGCATGAGGGCCGAGGTGTCGCCACGCGCCGCTCCAGTGGGC 64
11111111111111111111111111111111111111111111111111111
Db 370 GCCCACCAGCACTTGATGATGGGGGAGCCGGGTGTCACACACTTCTGCGCAGCAGCAGCAGC 311
11111111111111111111111111111111111111111111111111111
Cy 65 AGCGGTGAGGCGCATGAGCGGTGCGCCGCGCAGC 95
11111111111111111111111111111111111111111111111111111
Db 310 ATCACTGCTGTCATGCGCAGCAGCAGCAGC 280
11111111111111111111111111111111111111111111111111111
RESULT 14
AV391885/c      511 bp      mRNA      linear      EST 23-APR-2002
LOCUS
DEFINITION
Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
Accession
AV391885
Version
AV391885.1
Keywords
GI:6546101
Source
EST.
Organism
Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii.
Eukaryote, Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 511)
A Asamitsu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
A large scale structural analysis of cDNAs in a unicellular green
algae, Chlamydomonas reinhardtii. I. Generation of 3453
non-redundant expressed sequence tags
DNA Res. 6 (6), 369-373 (1999)
2015298
Journal
MEDLINE
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 511
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM081e12.r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      81 a      149 c      181 g      100 t
ORIGIN
Query Match      31.5%; Score 31.8; DB 10; Length 511;
Best Local Similarity 59.3%; Pred. No. 77;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
5 GCAAATCGCCCGCATGAGGGCCGAGGTGTCGCCACGCGCCGCTCCAGTGGGC 64
11111111111111111111111111111111111111111111111111111
Db 357 GCCCACCAGCACTTGATGATGGGGGAGCCGGGTGTCACACACTTCTGCGCAGCAGCAGCAGC 298
11111111111111111111111111111111111111111111111111111
Cy 65 AGCGGTGAGGCGCATGAGCGGTGCGCCGCGCAGC 95
11111111111111111111111111111111111111111111111111111

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900

Perfect score: 101
Sequence: 1 cggtcgaatcgccgcatg.....gctgcgcgcgcacgcacatc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PC1_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	36.6	11220	10	US-09-861-289-32
2	37	36.6	36778	10	US-09-861-289-5
3	29.8	29.5	1642	10	US-09-765-231A-25
4	28.4	28.1	2640	10	US-09-917-800A-1726
5	28.2	27.9	13058	10	US-09-764-846-303
6	27.6	27.3	2494	9	US-09-833-381-1984
7	27.4	27.1	4908	9	US-10-001-887-33
8	26.5	26.5	4661	10	US-09-285-385C-3
9	26	25.7	342	10	US-09-983-965-684
10	26	25.7	349	10	US-09-960-352-71
11	26	25.7	372	10	US-09-983-965-2663
12	26	25.7	873	10	US-09-891-718-3
13	26	25.7	1716	10	US-09-919-781-3
14	26	25.7	1917	10	US-09-919-781-1
15	26	25.7	2424	10	US-09-742-312-1
16	26	25.7	2724	10	US-09-891-718-1
17	26	25.7	2787	10	US-09-939-408A-16
18	26	25.7	147309	10	US-09-742-312-3
19	25.8	25.5	6000	10	US-09-764-860-768

c	20	25.8	25.5	43058	10	US-09-954-456-292	Sequence 292, App
c	21	25.8	25.5	43058	10	US-09-954-456-529	Sequence 529, App
c	22	25.8	25.5	43058	10	US-09-880-107-3950	Sequence 3950, App
c	23	25.4	25.1	373	10	US-09-864-761-21426	Sequence 21426, App
c	24	25.4	25.1	659	10	US-09-815-242-3978	Sequence 3978, App
c	25	25.2	25.0	1352	10	US-09-896-856-2	Sequence 2, App1
c	26	25.2	25.0	2430	10	US-09-861-289-23	Sequence 23, App1
c	27	25.2	25.0	3089	10	US-09-989-722-46	Sequence 46, App1
c	28	25.2	25.0	3089	10	US-09-988-723-46	Sequence 46, App1
c	29	25.2	25.0	3089	10	US-09-989-727-46	Sequence 46, App1
c	30	25.2	25.0	3089	10	US-09-989-731-46	Sequence 46, App1
c	31	25.2	25.0	3089	10	US-09-989-732-46	Sequence 46, App1
c	32	25.2	25.0	3089	10	US-09-990-442-46	Sequence 46, App1
c	33	25.2	25.0	3089	10	US-09-991-163-46	Sequence 46, App1
c	34	25.2	25.0	3089	10	US-09-993-604-46	Sequence 46, App1
c	35	25.2	25.0	3089	10	US-09-990-456-46	Sequence 46, App1
c	36	25.2	25.0	3089	10	US-09-989-721-46	Sequence 46, App1
c	37	25.2	25.0	3089	10	US-09-989-721-46	Sequence 46, App1
c	38	25.2	25.0	3089	10	US-09-989-721-46	Sequence 46, App1
c	39	25.2	25.0	3089	12	US-10-052-586-61	Sequence 61, App1
c	40	25.2	25.0	13613	10	US-09-861-289-3	Sequence 3, App1
c	41	25	24.8	534	10	US-09-738-973-265	Sequence 265, App
c	42	25	24.8	894	10	US-09-815-242-4029	Sequence 4029, App
c	43	25	24.8	1464	10	US-09-815-242-7915	Sequence 7915, App
c	44	25	24.8	1503	10	US-09-815-242-7680	Sequence 7680, App
c	45	25	24.8	1590	12	US-10-042-417-43	Sequence 43, App1

ALIGNMENTS

```
RESULT 1
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIORITY FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIORITY FILING DATE: 1996-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match          36.6%; Score 37; DB 10; Length 11220;
Best Local Similarity 60.4%; Pred. No. 0.014; Mismatches 40; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1 CGGTGCAATTCGCCGATGATGAGGGCCGAGTGTTCGCCAGCGAGCCGTCACAGT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9314 CCGTGACGCTGCCCGCGACTGGGGGTGGAGGTCCACGCGAGTACAGGGAAGT 9373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GGGCAGCGTTTCAGGCGATGGCGCGCGCGACGACATC 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 9374 GGGACGCCCTGCGCGGTGCGCTGGACGACGACGACATC 9414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
```

```
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding mathymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/361,269
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,517
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match 36.6%; Score 37; DB 10; Length 36778;
Best Local Similarity 60.4%; Pred. No. 0.014;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
```

```
QY 1 CGGTGCAATGCGCCGCGATGATAGGGCGAGGTGTTCGCCGCGAGCCCGTCAAGT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25001 CCGTGCAGCTGCGCCGCGACTGCGTGGAGTTCACGCGCAGCGAGTCAAGT 25060
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GGCAGCGGTTCAGCGCCATGCGCGTGC 36GACGCGACATC 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25061 GGCAGCGCTGCGCGCGCTGCGCTGGAGACGCGACATC 25101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 3

```
US-09-765-231A-25/C
; Sequence 25; Application US/09765231A
; Patent No. US20020119452A1
; GENERAL INFORMATION:
; APPLICANT: Searle/Monsanto
; APPLICANT: Philippe Desobry
; APPLICANT: Vasanthakumari, Geetha
; APPLICANT: Dolson, Stanton
; APPLICANT: Ma, Xiao-Jun
; TITLE OF INVENTION: Osteoarthritis disease-derived nucleic acids, polypeptides,
; FILE REFERENCE: SO-3221 PR
; CURRENT APPLICATION NUMBER: US/09/765,231A
; CURRENT FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 82
; SEQ ID NO 25
; LENGTH: 1642
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-765-231A-25
```

```
Query Match 29.5%; Score 29; DB 10; Length 1642;
Best Local Similarity 58.4%; Pred. No. 1.5;
Matches 52; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 13 CCCGATGATAGGGCGCCGAGTGTTCGCCAGGGGAGCCCGTCAAGTGGGAGCGGTC 72
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1476 CCCGATGATAGGGCGCCGAGTGTTCGCCAGGGGAGCCCGTCAAGTGGGAGCGGTC 1417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 73 AGGCGATGGCGGTGCGCGCGACGACATC 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 TGGCGGTGAGACCCGCGCGCGCTTAGC 1388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 4

```
US-09-917-800A-1726/C
; Sequence 1726; Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
```

```
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/230,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/230,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/232,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1726
; LENGTH: 2640
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_024163
US-09-917-800A-1726
```

```
Query Match 28.1%; Score 28; DB 10; Length 2640;
Best Local Similarity 56.4%; Pred. No. 4;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 8 AATGCCCGATGATAGGGCGCCGAGTGTTCGCCGCGAGCCCGTCAAGTGGGCGAGC 67
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 278 AATTCGAGATGATAGGGCGCCGAGTGTTCGCCGAGTGTTCGAGTGTTCGAGTGTTCG 219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 68 GGTTCAGCGCCATGCGCGTGGCGCGCGACGACATC 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 GGTTCAGCGCGCGTGGCGCGAGTGTTCGCCGAGTGTTCGCCGAGTGTTCGCCGAGTGTTC 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 5

```
US-09-764-846-303
; Sequence 303; Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 303
; LENGTH: 13058
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-846-303
```

```
Query Match 27.9%; Score 28; DB 10; Length 13058;
Best Local Similarity 61.6%; Pred. No. 4.8;
Matches 45; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
```

```
QY 4 TGCATATCGCCCGATGATAGGGCGCGAGTGTTCGCCGAGCGGCGCGCTCCCAAGTGGG 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


	Query Match	27.1%	Score 27.4	DB 9	Length 4908
Best Local Similarity	59.7%	Pred. No. 7.9			
Matches 46	Conservative	0	Mismatches 31	Indels 0	Gaps 0
QY	15	CGATGATAGGGCCGAGGTTCGACCCGCGACCCCTCCAACTGGGCAACGGTTAC	74		

```

RESULT 9
; Sequence 3684, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3684
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 43-LIB3058-043-Q1-K1-C12
; US-09-983-965-3684

```

Query Match 25.7%; Score 26; DB 10; Length 342;
Best Local Similarity 59.5%; Pred. No. 18;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 19 GGAATAGGGCGGAGGTTCGCCACAGCCGCTCCAGTGGCGAGCGGTTACAGGCCA 78
DB 182 GGGGCGGGGACAGCTGGCTGCCAGGCGCCGCGGTCCAGCGCGCGCGAGAGGCGC 123
QY 79 TGGCGCTGCCCGCC 92
DB 122 CGGCGAGCTGCGC 109

RESULT 10
US-09-960-352-71/c.
Sequence 71, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(102)8)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 71
LENGTH: 349
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 01-LIB3057-002-Q1-K1-A9
US-09-960-352-71

Query Match 25.7%; Score 26; DB 10; Length 349;
Best Local Similarity 59.5%; Pred. No. 18;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 19 GGAATAGGGCGGAGGTTCGCCACAGCCGCTCCAGTGGCGAGCGGTTACAGGCCA 78
DB 259 GGGGCGGGGACAGCTGGCTGCCAGGCGCCGCGGTCCAGCGCGCGAGAGGCGC 200
QY 79 TGGCGCTGCCCGCC 92
DB 199 CGGCGAGCTGCGC 186

RESULT 11
US-09-983-965-2663/c
Sequence 2663, Application US/09983963
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(102)9)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,131
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,478
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 2663
LENGTH: 372
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:

OTHER INFORMATION: Clone ID: 10-LIB3058-043-Q1-K1-C5
US-09-983-965-2663

Query Match 25.7%; Score 26; DB 10; Length 372;
Best Local Similarity 59.5%; Pred. No. 18;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 19 GGAATAGGGCGGAGGTTCGCCACAGCCGCTCCAGTGGCGAGCGGTTACAGGCCA 78
DB 149 GGGGCGGGGACAGCTGGCTGCCAGGCGCCGCGGTCCAGCGCGCGAGAGGCGC 90
QY 79 TGGCGCTGCCCGCC 92
DB 89 CGGCGAGCTGCGC 76

RESULT 12
US-09-891-718-3/c
Sequence 3, Application US/09891718
Patent No. US20020127671A1
GENERAL INFORMATION:
APPLICANT: Curtis, Rory A. J.
APPLICANT: Silos-Santiago, Imaculada
TITLE OF INVENTION: 52927, A NOVEL HUMAN CALCIUM CHANNEL AND
FILE REFERENCE: MNI-168
CURRENT APPLICATION NUMBER: US/09/891,718
CURRENT FILING DATE: 2001-06-26
PRIOR APPLICATION NUMBER: 60/214,101
PRIOR FILING DATE: 2000-06-26
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 873
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(873)
US-09-891-718-3

Query Match 25.7%; Score 26; DB 10; Length 873;
Best Local Similarity 62.1%; Pred. No. 19;
Matches 41; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 22 TAGGGCGGAGGTTCGCCACAGCGGCGTCCAGTGGCGAGCGGTTACAGGCCATGG 81
DB 500 TAGGCATAGAGTGTCTGCTGCGCGGCTCTCTGTTCCAGCGGGGTGAGCGTCAAGCGCTGG 441
QY 82 GCGTGC 87
DB 440 CCCTGC 435

RESULT 13
US-09-919-781-3/c
Sequence 3, Application US/09919781
Patent No. US20020125094A1
GENERAL INFORMATION:
APPLICANT: Curtis, Rory A. J.
TITLE OF INVENTION: 57250, A NOVEL HUMAN SUGAR TRANSPORTER
FILE REFERENCE: MNI-181
CURRENT APPLICATION NUMBER: US/09/919,781
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: USSN 60/221,769
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1716
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1716)
US-09-919-781-3

Query Match
Best Local Similarity 62.1%; Score 26; DB 10; Length 1716;
Matches 41; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 25 GGGCCGAGGTGTTGCCACGCGGCGGCGTCCAGTGGGCGAGCGGTTAGCGCATGGCG 84
DB 1664 GGCAGGTGCTCTGCGGGGCGGCCCGCAGAGTGGGGGCGCGCAGCGGTGCGCG 1605

QY 85 TGCCGC 90
DB 1604 TCCTGC 1599

RESULT 14
US-09-919-781-1/c
Sequence 1, Application US/09919781
Patent No. US20020123094A1
GENERAL INFORMATION:
APPLICANT: Curtis, Roy A.J.
TITLE OF INVENTION: 57250, A NOVEL HUMAN SUGAR TRANSPORTER
TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF
FILE REFERENCE: MNI-181
CURRENT APPLICATION NUMBER: US/09/919,781
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: USSN 60/221,769
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1917
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (13)...(1728)
US-09-919-781-1

Query Match
Best Local Similarity 62.1%; Score 26; DB 10; Length 1917;
Matches 41; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 25 GGGCCGAGGTGTTGCCACGCGGCGGCGTCCAGTGGGCGAGCGGTTAGCGCATGGCG 84
DB 1676 GGCAGGTGCTCTGCGGGGCGGCCCGCAGAGTGGGGGCGCGCAGCGGTGCGCG 1617

QY 85 TGCCGC 90
DB 1616 TCCTGC 1611

RESULT 15
US-09-742-312-1/c
Sequence 1, Application US/09742312
Patent No. US20020045166A1
GENERAL INFORMATION:
APPLICANT: CHANDRAMOULISARAN, Ishwar et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CLO00838
CURRENT APPLICATION NUMBER: US/09/742,312
CURRENT FILING DATE: 2000-12-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2424
TYPE: DNA
ORGANISM: Human

US-09-742-312-1

Query Match
Best Local Similarity 62.1%; Score 26; DB 10; Length 2424;
Matches 41; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 22 TAGGGCCGAGGTGTTGCCACGCGGCGGCGTCCAGTGGGCGAGCGGTTAGCGCATGG 81
DB 461 TAGCATAGAGTCTGCTGCGGCGCAGCTCTGTTCCACGGGCTGAGCGTCAAGCGCTGG 402

QY 82 GCGTGC 87
DB 401 CCCTGC 396

Search completed: November 5, 2002, 23:01:55
Job time: 94.2537 secs

;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9,1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 60
DB 35412 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 35471
OY 61 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 101
DB 35472 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 35512

RESULT 3
US-09-567-969-1
;; Sequence 1, Application US/09567969
;; Patent No. 6335457
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/567,969
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9,1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 60
DB 35412 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 35471
OY 61 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 101
DB 35472 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 35512

RESULT 4
US-09-568-480-1
;; Sequence 1, Application US/09568480
;; Patent No. 6335458
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,480
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9,1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 60
DB 35412 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 35471
OY 61 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 101
DB 35472 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 35512

RESULT 5
US-09-568-486-1
;; Sequence 1, Application US/09568486
;; Patent No. 6335459
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,486
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9,1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 60
DB 35412 CGGTGCAATCGCCCGATGATAGGGGCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 35471
OY 61 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 101
DB 35472 GGGCAGCGGTTTCAGGCGATGGCGGTGCGCGACGCACATC 35512

RESULT 6
US-09-568-472-1
;; Sequence 1, Application US/09568472
;; Patent No. 6338719
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,472
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9.1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGTGCAAAATGCGCCGATGATAGAGGGCCGAGGTGTCGCCACGGCGAGCCCGTCCAAGT 60
|||||
Db 35412 CGGTGCAAAATGCGCCGATGATAGAGGGCCGAGGTGTCGCCACGGCGAGCCCGTCCAAGT 35471
OY 61 GGGCAGCGGTTGAGGCGCATGGGCGCGCCGCGACGCACATC 101
|||||
Db 35472 GGGCAGCGGTTGAGGCGCATGGGCGCGCCGCGACGCACATC 35512

RESULT 7
US-09-567-899-1
Sequence 1, Application US/09567899
Patent No. 6383787
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,899
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9.1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGTGCAAAATGCGCCGATGATAGAGGGCCGAGGTGTCGCCACGGCGAGCCCGTCCAAGT 60
|||||
Db 35412 CGGTGCAAAATGCGCCGATGATAGAGGGCCGAGGTGTCGCCACGGCGAGCCCGTCCAAGT 35471
OY 61 GGGCAGCGGTTGAGGCGCATGGGCGCGCCGCGACGCACATC 101
|||||
Db 35472 GGGCAGCGGTTGAGGCGCATGGGCGCGCCGCGACGCACATC 35512

RESULT 8
US-09-443-501A-2
Sequence 2, Application US/09443501A
Patent No. 6303342
GENERAL INFORMATION:

APPLICANT: Kosan Biosciences, Inc.
APPLICANT: Julien, Bryan
APPLICANT: Katz, Leonard
APPLICANT: Khosla, Chaitan
APPLICANT: Tang, Li
APPLICANT: Ziermann, Rainer
TITLE OF INVENTION: Recombinant Methods and Materials for Producing
TITLE OF INVENTION: Epothilone and Epothilone Derivatives
FILE REFERENCE: 30062-20031.00
CURRENT APPLICATION NUMBER: US/09/443,501A
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: US 60/130,560
PRIOR FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: US 60/122,620
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/119,386
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: US 60/109,401
PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 71989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 9.1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGTGCAAAATGCGCCGATGATAGAGGGCCGAGGTGTCGCCACGGCGAGCCCGTCCAAGT 60
|||||
Db 29800 CGGTGCAAAATGCGCCGATGATAGAGGGCCGAGGTGTCGCCACGGCGAGCCCGTCCAAGT 29859
OY 61 GGGCAGCGGTTGAGGCGCATGGGCGCGCCGCGACGCACATC 101
|||||
Db 29860 GGGCAGCGGTTGAGGCGCATGGGCGCGCCGCGACGCACATC 29900

RESULT 9
US-08-804-227C-7
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
NAME/KEY: CDS
LOCATION: 14046..20036
NAME/KEY: CDS
LOCATION: 20110..31284
NAME/KEY: CDS
LOCATION: 31329..36071
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match
Best Local Similarity 49.3%; Score 49.8; DB 2; Length 44377;
Matches 69; Conservative 68.3%; Pred. No. 1.3e-05; Mismatches 32; Indels 0; Gaps 0;

QY 1 CGGTGCAATTCGCCGATGATAGAGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 60
DB 29396 CGGTGCAATTCGCCGATGATAGAGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 29455
QY 61 GGGCAGCGGTTCAGGCATGAGCGGCGTGCCTCCGACGACACATC 101
DB 29456 GGGCAGCGGTTCAGGCATGAGCGGCGTGCCTCCGACGACACATC 29496

RESULT 10
US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhlmann, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLATEINOLIDE SYNTHASE GENE
FILE REFERENCE: 24366-20007.00
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
CORRESPONDENCE ADDRESS:
ADDRESS: PAUL R. CANTRELL, 1118
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
```

```
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
NAME/KEY: CDS
LOCATION: 14046..20036
NAME/KEY: CDS
LOCATION: 20110..31284
NAME/KEY: CDS
LOCATION: 31329..36071
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match
Best Local Similarity 49.3%; Score 49.8; DB 2; Length 44377;
Matches 69; Conservative 68.3%; Pred. No. 1.3e-05; Mismatches 32; Indels 0; Gaps 0;

QY 1 CGGTGCAATTCGCCGATGATAGAGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 60
DB 29396 CGGTGCAATTCGCCGATGATAGAGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 29455
QY 61 GGGCAGCGGTTCAGGCATGAGCGGCGTGCCTCCGACGACACATC 101
DB 29456 GGGCAGCGGTTCAGGCATGAGCGGCGTGCCTCCGACGACACATC 29496

RESULT 11
US-09-103-840A-2/C
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match
Best Local Similarity 38.8%; Score 39.2; DB 4; Length 4403765;
Matches 62; Conservative 62.0%; Pred. No. 0.014; Mismatches 38; Indels 0; Gaps 0;

QY 1 CGGTGCAATTCGCCGATGATAGAGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAGT 60
DB 2304759 CGGTGCAATTCGCCGATGATAGAGGGCCGAGGTGTTCGCCACGGCGAGCGGAGGTAGT 2304700
QY 61 GGGCAGCGGTTCAGGCATGAGCGGCGTGCCTCCGACGACACATC 100
DB 29456 GGGCAGCGGTTCAGGCATGAGCGGCGTGCCTCCGACGACACATC 29496
```


Db 2304699 GGGACACCTTGCGCCATGGGCTTTGACGACGACACAT 2304660

RESULT 12

US-09-105-537-32

; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-32

Query Match

36.6%; Score 37; DB 4; Length 11220;

Best Local Similarity 60.4%; Pred. No. 0.04; Mismatches 40; Indels 0; Gaps 0;

Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 1 CGGTGCAATGCGCCGATGATGAGGGCGAGGTGTCGCGACGCGCCGTCCTCAAGT 60

Db 9314 CCGTGACAGCTGCGCGGCGCTGAGGTCCACGCGCGCGAGTCACGGGAAGT 9373

OY 61 GGGCAGCGGTCAGGCGCATGGGCGCGCGCGCGCGACGACATC 101

Db 9374 GGGACGCCCTGCGCGCTGCGCTGGACGACGCGGACATC 9414

RESULT 13

US-09-105-537-5

; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match

36.6%; Score 37; DB 4; Length 36778;

Best Local Similarity 60.4%; Pred. No. 0.043; Mismatches 40; Indels 0; Gaps 0;

Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 1 CGGTGCAATGCGCCGATGATGAGGGCGCGAGGTGTCGCGACGCGCCGTCCTCAAGT 60

Db 25001 CCGTGACAGCTGCGCGGCGCTGAGGTCCACGCGCGCGAGTCACGGGAAGT 25060

OY 61 GGGCAGCGGTCAGGCGCATGGGCGCGCGCGCGACGACATC 101

Db 25061 GGGACGCCCTGCGCGCTGCGCTGGACGACGCGGACATC 25101

RESULT 14

US-09-320-878-19

; Sequence 19, Application US/09320878A

; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match

36.6%; Score 37; DB 3; Length 38506;

Best Local Similarity 60.4%; Pred. No. 0.043; Mismatches 40; Indels 0; Gaps 0;

Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 1 CGGTGCAATGCGCCGATGATGAGGGCGCGAGGTGTCGCGACGCGCCGTCCTCAAGT 60

Db 23143 CCGTGACAGCTGCGCGGCGCTGAGGTCCACGCGCGCGAGTCACGGGAAGT 23202

OY 61 GGGCAGCGGTCAGGCGCATGGGCGCGCGCGCGCGACGACATC 101

Db 23203 GGGACGCCCTGCGCGCTGCGCTGGACGACGCGGACATC 23243

RESULT 15

US-07-642-734C-3

; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Edward R. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Dancigers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-938-2623
TELEFAX: 708-937-9396
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722 /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 19..4470 /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 97..1482 /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1693..2670 /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 3406..3921 /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4171..4428 /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4471..10722 /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 4471..5847 /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 6054..7026 /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 7165..9216 /function= "approximate span of
OTHER INFORMATION: dehydrogenase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc-feature

LOCATION: 9433..9984 /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 10225..10483 /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 10723..15165 /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 10831..12174 /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 12379..13350 /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 14062..14610 /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 14857..15114 /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 15166..20235 /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 15172..16569 /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 16768..17721 /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 18376..18921 /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 19149..19398 /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 19492..20235 /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

Query Match 35.6%; Score 36; DB 1; Length 20235;

Best Local Similarity	62.0%	Pred. No. 0.078;			
Matches	57; Conservative	0; Mismatches	35; Indels	0; Gaps	0;

10 TCGCCGATGATAGGGGCGAGGTTCGCCACGGCGAGCCGTCACAGTGGGCAACGG 69

Db 8948 TGGCCCCCGAGCGGGGGCGGAGTTTGGCCACCGCGGCCCGGCCCAAGCAGGGAGGC 9007

70 TTCAGGCCATGGCGTGCCCGCAGCACATC 101

Db 9008 TGGGGCGCTCGGTCTCGACGACGACATC 9039

```
Search completed: November 5, 2002, 16:03:55
Job time : 1239.74 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900

Perfect score: 101
Sequence: 1 cggtgcgaatccgccgcatg.....gcgtgcgcgcgcacgcacatc 101

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq-101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	68750	21	AAZ55887
2	101	100.0	71989	21	AAA29349
3	49.8	49.3	44377	18	AAT78508
4	49.8	49.3	44377	18	AAT80414
5	49.8	49.3	125401	22	AAAD17186
6	46.6	46.1	65140	22	AAAD17184
7	40.8	40.4	47981	22	AAF30757
8	39.2	38.8	4403765	22	AA199683
9	37	36.6	11220	21	AAZ87298

10	37	36.6	36778	21	AAZ87318
11	37	36.6	37948	21	AAZ87285
12	37	36.6	38506	21	AAZ75633
13	37	36.6	38506	21	AAZ56001
14	37	36.6	77536	21	AAA14651
15	36	35.6	29879	14	AAQ46806
16	35.6	35.2	4851	22	AAH52062
17	35.6	35.2	6459	22	AAF88336
18	35.6	35.2	43280	18	AAT80413
19	35.6	35.2	50000	22	AAF88316
20	35.6	35.2	50000	22	AAF88316
21	35.6	35.2	80161	20	AAZ21501
22	35.4	35.0	28598	17	AAT06769
23	35.4	35.0	28958	18	AAT89956
24	35.4	35.0	28958	18	AAV5289
25	35.4	35.0	49377	19	AAV05287
26	35.4	35.0	50937	21	AAV09469
27	32.8	32.5	30690	21	AAA92301
28	32.8	32.5	30690	22	AAH79277
29	32.6	32.3	352	18	AAT91483
30	32.6	32.3	352	18	AAT91420
31	32.6	32.3	352	19	AAV64469
32	32.6	32.3	352	19	AAV44361
33	32.6	32.3	352	20	AAZ19271
34	32.6	32.3	352	20	AAZ19059
35	30.6	30.3	77536	21	AAA14651
36	30.2	29.9	1263	22	AAAD08879
37	30	29.7	909	21	AAA11259
38	30	29.7	1365	24	ABO90174
39	29.8	29.5	1642	22	AAH23095
40	29.2	28.9	1985	21	AAZ36421
41	28.2	28.9	9359	22	AAK73225
42	28.8	28.5	1636	22	AAF45121
43	28.6	28.3	6375	23	ABL14967
44	28.6	28.3	8792	23	ABL14966
45	28.6	28.3	124884	22	AAH74201

ALIGNMENTS

RESULT 1	AAZ55887	standard; DNA; 68750 BP.
ID	AAZ55887	
XX		
AC	AAZ55887;	
XX		
DT	10-Apr-2000	(first entry)
DE	Sorangium cellulosum 68.75 kb contig.	
XX		
KW	Epithione biosynthesis; type I polyketide synthase; taxol substitute; anticancer; ds.	
KW		
XX		
OS	Sorangium cellulosum.	
XX		
FH	key	Location/Qualifiers
FT	CDS	1..1826
FT		/*tag= a
FT		/partial
FT		/product= "Partial Orf 1 protein (AAV58580)"
FT		/note= "No initiation codon given in the specification"
FT		complement (1900..3171)
FT	CDS	/*tag= b
FT		/product= "Orf 2 protein (AAV58581)"
FT		3415..5556
FT	CDS	/*tag= c
FT		/product= "Orf 3 protein (AAV58582)"
FT		complement (5612..5992)
FT	CDS	/*tag= d
FT		/product= "Orf 4 protein (AAV58583)"
FT		6226..6675
FT	CDS	/*tag= e

S. venezuelae pik
S. venezuelae pik
Nucleotide sequenc
Recombinant cosmid
Nucleotide sequenc
eryA region of S.
S. spinosa DNA tube
S. spinosa DNA fra
Tylactone synthase
S. spinosa DNA fra
S. spinosa DNA fra
DNA fragment of Sa
Sorangium cellulos
Sorangium cellulos
DNA sequence of So
The soraphen biosy
Streptococcus olea
S. avermitilis ave
Streptococcus aveym
Mycobacterium tube
Mycobacterium tube
Mycobacterium tube
M. tuberculosis Im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Nucleotide sequenc
Mycobacterium szul
Burkholderia strai
M. capsulatus gene
Osteoarthritis tis
DNA encoding a mai
Human immune/haema
Human ITAGO 202 cd
Drosophila melanog
Drosophila melanog
Nucleotide sequenc

```

FT      CDS      /product= "Orf 5 protein (AAV58584)"
FT      FT      /tag- f
FT      CDS      /product= "Type j polypeptide synthase, EPOS A
FT      FT      (AAV51573)"
FT      CDS      11872..116104
FT      FT      /tag- g
FT      CDS      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      FT      (AAV58574)"
FT      CDS      16251..21749
FT      FT      /tag- h
FT      CDS      /product= "Type i polypeptide synthase, EPOS B
FT      FT      (AAV58575)"
FT      CDS      21746..43519
FT      FT      /tag- i
FT      CDS      /product= "Type i polypeptide synthase, EPOS C
FT      FT      (AAV58576)"
FT      CDS      43524..54920
FT      FT      /tag- j
FT      CDS      /product= "Type i polypeptide synthase, EPOS D
FT      FT      (AAV58577)"
FT      CDS      54935..62254
FT      FT      /tag- k
FT      CDS      /product= "Type i polypeptide synthase, EPOS E
FT      FT      (AAV58578)"
FT      CDS      62369..63628
FT      FT      /tag- l
FT      CDS      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      FT      (AAV58579)"
FT      CDS      63779..64333
FT      FT      /tag- m
FT      CDS      /product= "Orf 6 protein (AAV58585)"
FT      FT      complement (63853..64290)
FT      CDS      /tag- n
FT      CDS      /product= "Orf 7 protein (AAV58586)"
FT      FT      64363..64920
FT      CDS      /tag- o
FT      CDS      /product= "Orf 8 protein (AAV58587)"
FT      FT      complement (64287..64727)
FT      CDS      /tag- p
FT      CDS      /product= "Orf 9 protein (AAV58588)"
FT      FT      65063..65767
FT      CDS      /tag- q
FT      CDS      /product= "Orf 10 protein"
FT      FT      complement (65008..65874)
FT      CDS      /tag- r
FT      CDS      /product= "Orf 11 protein (AAV58590)"
FT      FT      complement (65871..66338)
FT      CDS      /tag- s
FT      CDS      /product= "Orf 12 protein (AAV58591)"
FT      FT      66667..67137
FT      CDS      /tag- t
FT      CDS      /product= "Orf 13 protein (AAV58592)"
FT      FT      67334..68251
FT      CDS      /tag- u
FT      CDS      /product= "Orf 14 protein (AAV58593)"
FT      FT      68346..68750
FT      CDS      /partial
FT      CDS      /tag- v
FT      CDS      /product= "Partial Orf 15 protein (AAV58594)"
FT      FT      /note= "No termination codon given in the specification
XX      MO9966028-A2.
XX
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
XX      05-FEB-1999; 99US-0118906.
XX      (NOVS ) NOVARTIS AG.

```

XX	PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX	PI	Schupp T., Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX	DR	WPI: 2000-097741/08.
XX	DR	P-PSDB: AAY58572, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578.
XX	DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584, AAY58585, AAY58586, AAY58587, AAY58588, AAY58589, AAY58590, AAY58591, AAY58592, AAY58593, AAY58594.
XX	DR	
XX	PT	New isolated epoethione synthase genes, used for the recombinant production of epoethione for use in cancer therapy -
XX	PT	
XX	PT	Claim 14; Page 87-104; 174pp: English.
XX	PS	
XX	XX	This sequence represents a 68.75 kb contig from Sorangium cellulosum comprising 22 open reading frames (ORFs) and includes genes encoding proteins involved in the biosynthesis of epoethiones. Epoethiones A and B are 16-membered macrocyclic polyketides with an acylcysteine-derived starter unit; polyketides being synthesised from two-carbon building blocks, the beta-carbon of which always carries a keto group. Each round of two-carbon addition is carried out by a complex of enzymes known as the polyketide synthase in a manner similar to fatty acid biosynthesists. EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D and EPOS E (AAY58575-Y58578) are involved in polyketide backbone formation. EPO F (AAY58579) is an epoethione macrolactone oxidase, and the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be involved in transport. Epoethiones mimic the biological activity of taxol, and may be substituted for taxol in cancer chemotherapeutic compositions. Epoethiones exhibit a much lower drop in potency against a multiply drug-resistant cell line compared with taxol, by the multidrug considerably less efficiently exported from such cells, and are resistance protein (MDR, or P-glycoprotein). Despite the potential of epoethiones as anticancer agents, they are problematical to produce on a large scale. Epoethiones are too complex for industrial scale chemical synthesis and Sorangium cellulosum is difficult to ferment producing poor yields of epoethiones. The nucleic acids of the invention may be used for the recombinant production of epoethiones in a heterologous host that is more amenable to fermentation.
XX	XX	Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX	XX	Query Match 100.0%; Score 101; DB 21; Length 68750;
XX	XX	Best Local Similarity 100.0%; Pred. No. 1.2e-18;
XX	XX	Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1	CGGTGCAAAATGCGCCGATGATAGGAGGCGCGAGGTGTCGCCAGCGACCGCTTCCAAAGT 60
DB	35412	CGGTGCAAAATGCGCCGATGATAGGAGGCGCGAGGTGTCGCCAGCGACCGCTTCCAAAGT 35471
QY	61	GGGCGAGCGGTTTCAGGCGCATGGGCGGTGCGCGCGACGACATC 101
DB	35472	GGGCGAGCGGTTTCAGGCGCATGGGCGGTGCGCGCGACGACATC 35512
XX	XX	Result 2
XX	XX	AAA29349
XX	XX	ID AAA29349 standard; DNA: 71989 BP.
XX	XX	AAA29349;
XX	XX	12-SEP-2000 (first entry)
XX	XX	Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX	XX	Epoethione: polyketide synthase; epoa, epob, epoc, epod, epoe, epof, epog, epok, P450 epoxidase; ORFa, ORFb, promoter; enhancer; anti-fungal; tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX	XX	Sorangium cellulosum.
XX	XX	Key Location/Qualifiers
XX	XX	

```

FT      CDS      3..992      /*tag= a
FT      /label= ORF A
FT      /product= transposase
FT      /note= "not part of the PKS"
FT      989..1501
FT      /*tag= b
FT      /label= ORF-B
FT      /product= transposase
FT      /note= "not part of the PKS"
FT      1998..6263
FT      CDS      /*tag= c
FT      /label= epaA_gene
FT      /note= "encodes the loading domain"
FT      2031..3548
FT      /*tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      3621..4661
FT      /*tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      4917..5810
FT      /*tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      5856..6155
FT      /*tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      6260..10493
FT      /*tag= h
FT      /label= epoB_gene
FT      /note= "encodes module 1, the NRPS module"
FT      2031..3548
FT      /*tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      2031..3548
FT      /*tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      6861..6887
FT      /*tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      6861..6887
FT      /*tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      7358..7386
FT      /*tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      7898..7921
FT      /*tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      7898..7921
FT      /*tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      8261..8308
FT      /*tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      8411..8422
FT      /*tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      8861..8905
FT      /*tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      8966..8983
FT      /*tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      9090..9179
FT      misc_RNA

FT      /*tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      9183..9992
FT      /*tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      10121..10138
FT      /*tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      10261..10306
FT      /*tag= w
FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      10639..16137
FT      /*tag= x
FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      10654..12033
FT      /*tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      12250..13287
FT      /*tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      13327..13899
FT      /*tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      14962..15756
FT      /*tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      15763..16008
FT      /*tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      16134..37907
FT      /*tag= ad
FT      /label= epod_gene
FT      /note= "encodes modules 3-6"
FT      16425..17606
FT      /*tag= ae
FT      /note= "encodes KS3"
FT      17817..18857
FT      /*tag= af
FT      /note= "encodes AT3"
FT      19581..20396
FT      /*tag= ag
FT      /note= "encodes KR3"
FT      20424..20642
FT      /*tag= ah
FT      /note= "encodes ACP3"
FT      20706..22082
FT      /*tag= ai
FT      /note= "encodes KS4"
FT      22296..23336
FT      /*tag= aj
FT      /note= "encodes AT4"
FT      24069..24647
FT      /*tag= ak
FT      /note= "encodes KR4"
FT      24867..25151
FT      /*tag= al
FT      /note= "encodes ACP4"
FT      25203..26576
FT      /*tag= am
FT      /note= "encodes KS5"
FT      26793..27883
FT      /*tag= an
FT      /note= "encodes AT5"
FT      27966..28574
FT      /*tag= ao
FT      /note= "encodes DH5"
FT      29433..30287
FT      /*tag= ap
FT      misc_RNA

```

```
FT      misc_RNA      /note= "encodes HR5"
FT      30321..30869
FT      /tag= aq
FT      /note= "encodes KR5"
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      31440..32807
FT      /tag= as
FT      /note= "encodes KS6"
FT      33018..34067
FT      /tag= at
FT      /note= "encodes AR6"
FT      34107..34676
FT      /tag= au
FT      /note= "encodes DI6"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes EI6"
FT      36705..37256
FT      /tag= aw
FT      /note= "encodes KI6"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= epep gene
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT7"
FT      41341..41922
FT      /tag= bb
FT      /note= "encodes KF7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.2e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 CGGTCAATCGCCCGATGATAGGGCCGAGGTGTTCCGACGCGGAGCCCGTCCACT 60
DB      29800 CGGTCAATCGCCCGATGATAGGGCCGAGGTGTTCCGACGCGGAGCCCGTCCACT 29859

OY      61 GGCAGCGGTTTCAGGCATGCGCGTCCGCGCACGCACATC 101
DB      29860 GGCAGCGGTTTCAGGCATGCGCGTCCGCGCACGCACATC 29900

RESULT 3
ID      AAT78508 standard; DNA: 44377 BP.
AC      AAT78508;
DT      26-FEB-1998 (first entry)
DE      Platenolide synthase gene cluster.
XX      Platenolide synthase gene cluster.
KW      platenolide synthase gene cluster; platenolide production; srmg gene;
KW      multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX      Streptomyces ambofaciens.
XX
XX      Key      Location/Qualifiers
FH      CDS      350..14002
FT      /tag= a
FT      /transl_except= (pos:350..352, aa:Met)
FT      /note= "ORF1 encodes protein shown in AAW23716"
```

```
FT      CDS      14046..20036
FT      /tag= b
FT      /note= "ORF2 encodes protein shown in AAW23717"
FT      20110..31284
FT      /tag= c
FT      /transl_except= (pos:20111..20113, aa:Met)
FT      /note= "ORF3 encodes protein shown in AAW23718"
FT      31329..36071
FT      /tag= d
FT      /note= "ORF4 encodes protein shown in AAW23719"
FT      36155..41830
FT      /tag= e
FT      /note= "ORF5 encodes protein shown in AAW23720"

XX      EP791656-A2.
XX      27-AUG-1997.
XX      19-FEB-1997; 97EP-0301066.
XX      22-FEB-1996; 96US-0012050.
XX      (E.L.I.L.) LILLY & CO E.L.I.
XX      Burgett SG, Kuhstoss SA, Rao RN, Richardson MA:
XX      Rostleek PR.
XX      WPI; 1997-418047/39.
XX      P-PSDB; AAW23716-W23720.
XX      DNA encoding Streptomyces ambofaciens platenolide synthase domain -
XX      for production of spiramycin-related polyketide antibiotics
XX      Claim 9; Pages 8-33; 81pp; English.
XX
XX      This sequence represents the platenolide synthase gene cluster of the
XX      invention. This sequence is referred to as the srmg gene, and was
XX      isolated from Streptomyces ambofaciens. This sequence encodes the
XX      multi-functional proteins which direct the synthesis of the polyketide
XX      platenolide. Platenolide is the basic building block of the macrolide
XX      antibiotic spiramycin. The DNA can be used to produce compounds
XX      exhibiting antibiotic activity based on the platenolide structure,
XX      including specifically the macrolide antibiotic spiramycin and spiramycin
XX      analogues and derivatives. Modifications of the platenolide synthase DNA
XX      sequence can be made so as to change the number and type of carboxylic
XX      acids incorporated into the growing polyketide chain and to change the
XX      kind of post-condensation processing that is conducted.
XX
XX      Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

Query Match      49.3%; Score 49.8; DB 18; Length 44377;
Best Local Similarity 68.3%; Pred. No. 0.00011;
Matches 69; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY      1 CGGTCAATCGCCCGATGATAGGGCCGAGGTGTTCCGACGCGGAGCCCGTCCACT 60
DB      29396 CGGTCAATCGCCCGATGATAGGGCCGAGGTGTTCCGACGCGGAGCCCGTCCACT 29455

OY      61 GGCAGCGGTTTCAGGCATGCGCGTCCGCGCACGCACATC 101
DB      29456 GGCAGCGGTTTCAGGCATGCGCGTCCGCGCACGCACATC 29496

RESULT 4
ID      AAT80414 standard; DNA: 44377 BP.
AC      AAT80414;
DT      27-FEB-1998 (first entry)
DE      Platenolide synthase gene cluster.
XX
```


DR WPI: 2001-557614/62.
DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX AAE10149, AAE10150.
PT New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
PS Claim 1; Page 188-254; 266pp; English.
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
Query Match 49.3%; Score 49.8; DB 22; Length 125401;
Best Local Similarity 68.3%; Pred. No. 0.00011;
Matches 69; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
OY 1 CGGTCAATCCCGCATGATAGGGGCGAGGTGTCGCCAGCGGCCGTCCTCAAGT 60
DB 39503 CGATCCAGATCCCGCGCATGTCGGTGCAGAGTGTCCGACGCGCATGAGGCCAAGT 39562
OY 61 GGGCAGCGGTTGAGGCGATGGCGTCCGCGCAGCAGCAGCATC 101
DB 39563 GGGACGTGTGGCGTCCCTGGCGGTGGCGCAGCAGCAGCATC 39603
RESULT 6
AADI7184
ID AADI7184 standard; DNA; 65140 BP.
XX AADI7184;
AC AADI7184;
DT 29-NOV-2001 (first entry)
XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nys1; ds.
XX Streptomyces noursei.
OS
XX Location/Qualifiers
FH complement (1..1033)
FT /*tag= a
FT /product= "NysD2 partial protein"
FT /note= "CDS does not include stop codon"
FT complement (1056..3576)
FT /*tag= b
FT /product= "NysD1 protein"
FT 2806..6906
FT /*tag= c
FT /product= "NysA protein"
FT 6952..16530
FT /*tag= d
FT /product= "NysB protein"
FT 16530..49840.
FT /*tag= e
FT /product= "NysC protein"
FT 50260..51015
FT /*tag= f
FT /product= "NysE protein"
FT 51405..54305
FT /*tag= g
FT /product= "NysR1 protein"
FT 54329..57190
FT /*tag= h
FT /product= "NysR2 protein"
FT /note= "CDS does not include start codon"
FT 57180..59963
CDS

FT /*tag= i
FT /product= "NysR3 protein"
FT 60415..61047
FT /*tag= j
FT /product= "NysR4 (short) protein"
FT /note= "CDS does not include start codon"
FT 61736..62497
FT /*tag= k
FT /product= "NysR5 protein"
FT /note= "CDS does not include start codon"
FT complement (62551..63615)
FT /*tag= l
FT /product= "ORF2 protein"
FT /note= "CDS does not include start codon"
FT 63765..64961
FT /*tag= m
FT /product= "ORF1 protein"
FT
FT CDS
FT
FT CDS
FT
FT WO200159126-A2.
FT
FT 16-AUG-2001.
FT
FT 08-FEB-2001; 2001WO-GB00509.
FT
FT 08-FEB-2000; 2000GB-0002840.
FT
FT 10-APR-2000; 2000GB-0008786.
FT
FT 14-APR-2000; 2000GB-0009387.
FT
FT (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
FT (SNTF) SINTER STIFTTELSEN IND TEK FORSK.
FT (ALPH-) ALPHARMA AS.
FT (SINV-) SINVENT AS.
FT (DZIE-) DZIELEWSKA H.
FT (ZOTC-) ZOTCHEV S.B.
FT (SEKU-) SEKUROVA O.N.
FT (FJAE-) FJAEVYK E.
FT (BRAU-) BRAUTASET T.
FT (STRO-) STROM A.R.
FT
FT Zolchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
FT Valla S, Ellingsen TE, Sletta H, Gulliksen O;
FT WPI: 2001-557614/62.
FT P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
PS Claim 2; Page 116-151; 266pp; English.
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
SQ Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
Query Match 46.1%; Score 46.6; DB 22; Length 65140;
Best Local Similarity 66.3%; Pred. No. 0.0008;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
OY 1 CGGTCAATCCCGCATGATAGGGGCGAGGTGTCGCCAGCGGCCGTCCTCAAGT 60
DB 31785 CGATCCAGATCCCGCGCATGTCGGTGCAGAGTGTCCGACGCGCATGAGGCCAAGT 31844
OY 61 GGGCAGCGGTTGAGGCGATGGCGTCCGCGCAGCAGCAGCATC 101
DB 31845 GGGACGTGTGGCGTCCCTGGCGGTGGCGCAGCAGCAGCATC 31885
RESULT 7

AAE30757	AAE30757 standard; DNA; 47981 BP.	FT	CDS	12181..22821
XX	AAF30757	FT		/gene= "mega"
AC	AAF30757;	FT		/tag= 1
XX	21-JUN-2001 (first entry)	FT		/product= "megalomiclin 6-deoxyerthrionolide B synthase 1"
DE	Micromonospora megalomiclin biosynthetic gene cluster.	FT	misc_feature	/note= "polyketide synthase; encodes AAB82212"
XX	Megalomiclin; meg gene; polyketide synthase; antibiotic; ds.	FT		/tag= m
XX		FT		/gene= "mega"
XX		FT		/function= "At-L"
OS	Micromonospora megalomicla subsp. nigra.	FT	misc_feature	/tag= n
XX		FT		/gene= "mega"
XX		FT		/function= "ACP-L"
XX		FT		/tag= o
XX		FT		/gene= "mega"
XX		FT		/function= "KSI"
XX		FT		/tag= p
XX		FT		/gene= "mega"
XX		FT		/function= "AT1"
XX		FT		/tag= q
XX		FT		/gene= "mega"
XX		FT		/function= "KRI"
XX		FT		/tag= r
XX		FT		/gene= "mega"
XX		FT		/function= "ACP1"
XX		FT		/tag= s
XX		FT		/gene= "mega"
XX		FT		/function= "KS2"
XX		FT		/tag= t
XX		FT		/gene= "mega"
XX		FT		/function= "AT2"
XX		FT		/tag= u
XX		FT		/gene= "mega"
XX		FT		/function= "KR2"
XX		FT		/tag= v
XX		FT		/gene= "mega"
XX		FT		/function= "ACP2"
XX		FT		/tag= w
XX		FT		/gene= "megAI1"
XX		FT		/product= "megalomiclin 6-deoxyerythronolide B synthase 2"
XX		FT		/note= "polyketide synthase, encodes AAB82213"
XX		FT		/tag= x
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS3"
XX		FT		/tag= y
XX		FT		/gene= "megAI1"
XX		FT		/function= "AT3"
XX		FT		/tag= z
XX		FT		/gene= "megAI1"
XX		FT		/function= "KR3 (inactive)"
XX		FT		/tag= aa
XX		FT		/gene= "megAI1"
XX		FT		/function= "AC3"
XX		FT		/tag= ab
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS4"
XX		FT		/tag= ac
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS5"
XX		FT		/tag= ad
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS6"
XX		FT		/tag= ae
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS7"
XX		FT		/tag= af
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS8"
XX		FT		/tag= ag
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS9"
XX		FT		/tag= ah
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS10"
XX		FT		/tag= ai
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS11"
XX		FT		/tag= aj
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS12"
XX		FT		/tag= ak
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS13"
XX		FT		/tag= al
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS14"
XX		FT		/tag= am
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS15"
XX		FT		/tag= an
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS16"
XX		FT		/tag= ao
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS17"
XX		FT		/tag= ap
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS18"
XX		FT		/tag= aq
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS19"
XX		FT		/tag= ar
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS20"
XX		FT		/tag= as
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS21"
XX		FT		/tag= at
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS22"
XX		FT		/tag= au
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS23"
XX		FT		/tag= av
XX		FT		/gene= "megAI1"
XX		FT		/function= "KS24"
XX		FT		/tag= aw
XX		FT		/gene= "megAI1"
XX		FT		

```
FT      /*tag= ac
FT      /gene= "megAII"
FT      /function= "AT4"
FT      29953..30477
FT      misc_feature
FT      /*tag= ad
FT      /gene= "megAII"
FT      /function= "DH4"
FT      31396..33244
FT      misc_feature
FT      /*tag= ae
FT      /gene= "megAII"
FT      /function= "ER4"
FT      32257..33299
FT      misc_feature
FT      /*tag= af
FT      /gene= "megAII"
FT      /function= "KR4"
FT      33052..33312
FT      misc_feature
FT      /*tag= ag
FT      /gene= "megAII"
FT      /function= "ACP4"
FT      33666..43271
FT      CDS
FT      /*tag= ah
FT      /gene= "megAII"
FT      /product= "megalomicin 6-deoxyerythronolide B
FT      synthase 3"
FT      /note= "polyketide synthase; encodes AAB82214"
FT      22957..24237
FT      misc_feature
FT      /*tag= ai
FT      /gene= "megAII"
FT      /function= "KS5"
FT      24544..25581
FT      misc_feature
FT      /*tag= aj
FT      /gene= "megAII"
FT      /function= "AT5"
FT      26230..26733
FT      misc_feature
FT      /*tag= ak
FT      /gene= "megAII"
FT      /function= "KR5"
FT      26998..27258
FT      misc_feature
FT      /*tag= al
FT      /gene= "megAII"
FT      /function= "ACPs"
FT      27393..28590
FT      misc_feature
FT      /*tag= am
FT      /gene= "megAII"
FT      /function= "KS6"
FT      28897..29931
FT      misc_feature
FT      /*tag= an
FT      /gene= "megAII"
FT      /function= "AT6"
FT      29953..30477
FT      misc_feature
FT      /*tag= ao
FT      /gene= "megAII"
FT      /function= "KR6"
FT      31396..33244
FT      misc_feature
FT      /*tag= ap

Query Match      40.4%; Score 40.8; DB 22: Length 47981;
Best Local Similarity 65.2%; Pred. No. 0.03; 32; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY      10 TCGCCCGATGATAGGCGCGAGTGTTCACAGCGGACCGCGTCCAGTAGTGGGACGG 69
DB      31772 TGGCCCGTGGGCGCGGCGAGTGTTCACAGCGGACCGCGCGCAACACCGCGC 31831
OY      70 TTCAGGCGCATGGCGCTGCCCGCACGCAATC 101
DB      31832 TGGCGGCGCTGGCGCTGCAGACGACCAATC 31863

RESULT 8
AAI99683/c
ID      AAI99683 standard; DNA: 4403765 BP.
XX
```

```
AC      AAI99683;
XX
XX      15-JAN-2002 (first entry)
DT
XX      Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
DE      Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome:
XX      variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX      Mycobacterium tuberculosis.
XX
XX      US6294328-B1.
XX
XX      25-SEP-2001.
XX
XX      24-JUN-1998; 98US-0103840.
XX
XX      24-JUN-1998; 98US-0103840.
XX
XX      (GENO-) INST GENOMIC RES.
XX
XX      Fleischmann RD, White OR, Fraser CM, Venter JC;
XX      WPI; 2001-647261/74.
XX
XX      Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX      determining the nucleotide sequence of the strain at positions in the
XX      genome corresponding to positions where M. tuberculosis strains CDC
XX      1551 and H37Rv differ -
XX
XX      Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
XX      The invention relates to evaluating strain variation within and between
XX      different populations of the tuberculosis bacterial pathogen,
XX      Mycobacterium tuberculosis or related Mycobacterium by determining the
XX      nucleotide sequence of the first strain at positions in the complete
XX      sequence of the genome that correspond to positions that differ in the
XX      nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
XX      H37Rv (AAI99682). The method is useful for evaluating strain variation of
XX      M. tuberculosis and has valuable application in the fields of
XX      tuberculosis genetics, epidemiology, patient treatment and epidemic
XX      monitoring.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from USPTO
XX      at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
XX      Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
XX
Query Match      38.8%; Score 39.2; DB 22: Length 4403765;
Best Local Similarity 62.0%; Pred. No. 0.096;
Matches 62; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

OY      1 CGGTGCAATTCGCGCATGATAGGCGCGAGTGTTCGCGACGCGCGTCCCAAGT 60
DB      2304759 CGGTGACAGTGTGCTGGCGATCGGGGTTGAGAGTGTTCGCGACCGCGCAAGGTAAGT 2304700
OY      61 GGGCAGCGGTTCAGGCGCATGAGGCGCGTCCCGCGACGCAAT 100
DB      2304699 GGGACACCTTGGCGCATGAGGCTTTGACGAGACCAAT 2304660

RESULT 9
AAI87298
ID      AAI87298 standard; DNA: 11220 BP.
XX
XX      AAI87298;
XX
XX      05-JUN-2000 (first entry)
DT
XX      S. venezuelae macrolide biosynthetic gene plkAII, SEQ ID NO:32.
XX      Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
XX      neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX      KW
```


CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polypeptides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC 15439, as given in figure 31.

SO Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other:

Query Match 36.6%; Score 37; DB 21; Length 36778;

Best Local Similarity 60.4%; Pred. No. 0.32;

Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 1 CGGTGCAATCGCCCGATGATGAGGGC/CGAGGTGTTCCGACGCGAGCCGTCACACT 60

DB 25001 CCGTCGACGCTCGCCCGCAGCTGGGCGCTGAGGTCCACGCGAGTACAGGGAAGT 25060

OY 61 GGGCAGCGGTTGAGCCATGCGGCTGCTCCGACGCAATC 101

DB 25061 GGGACGCCCTGCGCGCTGCTGAG/ACGCGCAATC 25101

RESULT 11

AAZ87285

ID AAZ87285 standard; DNA; 37948 BP.

AC AAZ87285;

XX 05-JUN-2000 (first entry)

DE S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.

KM Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;

KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;

KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;

KM chronic obstructive pulmonary disease; respiratory inflammation;

KM hypercholesterolaemia; crop protection agent; ds.

XX Streptomyces venezuelae ATCC15439.

XX WO200000620-A2.

PD 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MINU) UNIV MINNESOTA.

PI Sherman DH, Liu H, Xue Y, Zhao J;

XX WPI: 2000-160679/14.

DR P-PSDB; AA77180.

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

PT synthesis of methymycin and pikromycin -

XX Claim 13; Page 299-315; 438pp; Eng1sh.

XX The invention relates to an isolated and purified nucleic acid segment

CC comprising a desosamine biosynthetic gene cluster, a fragment or its

CC biologically active variant, where the nucleic acid sequence is not

CC derived from the ergC gene cluster of Saccharopolyspora erythraea or

CC Streptomyces antibiotics. The invention also relates to a macrolide

CC biosynthetic gene cluster, or fragments thereof. The macrolide

CC biosynthetic gene cluster encodes proteins which synthesise methymycin,

CC pikromycin, neomethymycin, narbomycin or a combination of these

CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrorides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyhydroxyalkanoate (PHA)
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polypeptides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.

SO Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other:

Query Match 36.6%; Score 37; DB 21; Length 37948;

Best Local Similarity 60.4%; Pred. No. 0.32;

Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 1 CGGTGCAATCGCCCGATGATGAGGGC/CGAGGTGTTCCGACGCGAGCCGTCACACT 60

DB 26171 CCGTCGACGCTCGCCCGCAGCTGGGCGCTGAGGTCCACGCGAGTACAGGGAAGT 26230

OY 61 GGGCAGCGGTTGAGCCATGCGGCTGCTCCGACGCAATC 101

DB 26231 GGGACGCCCTGCGCGCTGCTGAG/ACGCGCAATC 26271

RESULT 12

AAZ75633

ID AAZ75633 standard; DNA; 38506 BP.

AC AAZ75633;

XX 22-JAN-2001 (first entry)

DE Nucleotide sequence of the insert DNA in cosmid PKOS03-27.

KM Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;

KM antibiotic; C12-hydroxylase; pik; desosamine biosynthesis;

KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;

XX pikromycin biosynthesis; ss.

XX Streptomyces venezuelae.

XX US6117659-A.

XX 12-SEP-2000.

XX 27-MAY-1999; 99US-0320878.

XX 28-MAY-1998; 98US-0087080.

XX 22-SEP-1998; 98US-0100880.

XX 08-FEB-1999; 98US-0119139.

XX 30-MAY-1997; 97US-0134990.

XX 28-MAY-1996; 98US-0073538.

XX 28-AUG-1998; 98US-0141908.

XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;

XX WPI: 2000-610844/58.

PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
for converting ketolides to antibiotics and as antibiotics and

PT intermediates in the synthesis of compounds with pharmaceutical value
PT
XX
PS Disclosure; Columns 15-32; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxylase (pic), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 36.6%; Score 37; DB 21; Length 38506;
Best Local Similarity 60.4%; Pred. No. 0.33;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 1 CGGTGCAAAATGCCGATGATAGGGCCGAGTGTCCGACGCGGACCCGTCACAGT 60
DB 23143 CCGTGCACTGCGCCGACACTGGGGCGGAGGTCCACGCGGACGAGTCACGGGAAGT 23202

OY 61 GGGCAGCGGTTTCAGGCGCATGGCGCTGCCGCGCAGCAGCATATC 101
DB 23203 GGGACGCGCTCGCGGCGCTGGAGCAGCGGCGCATATC 23243

RESULT 13
AAZ56001
ID AAZ56001 standard; DNA; 38506 BP.
XX
AC AAZ56001;
XX
DT 23-MAR-2000 (first entry)
XX
DE Recombinant cosmid PKOS023-27 containing S. venezuelae FKS genes.
XX
KM Narbonolide polyketide synthase; PKS; cosmid PKOS023-27; ketolide;
KM antibiotic production; narbomycin; picromycin; ds.
XX
OS Streptomyces venezuelae.
XX
FH Location/Qualifiers
FT CDS 70..13725
FT /tag= a
FT /product= PICAI
FT /note= "Narbonolide synthase subunit 1"
FT 13830..25049
FT /tag= b
FT /product= PICAI1
FT /note= "Narbonolide synthase subunit 2"
FT 25133..29821
FT /tag= c
FT /product= PICAI11
FT /note= "Narbonolide synthase subunit 3"
FT 29924..33964
FT /tag= d
FT /product= PICAIIV
FT /note= "Narbonolide synthase subunit 4"
FT 33961..34806
FT /tag= e
FT /product= PICB
FT /note= "Contains typeII thioesterase domain"
FT 34863..36011
FT /tag= f

PT /product= PICCII
FT /note= "4-keto-6-deoxyglucose isomerase"
FT CDS 36159..37439
FT /tag= g
FT /product= PICCIII
FT /note= "Desosaminyl transferase"
FT CDS 37529..38242
FT /tag= h
FT /product= PICCVI
FT /note= "3-amino dimethyltransferase"
XX
XX W09961599-A2.
XX
XX 02-DEC-1999.
XX
XX 27-MAY-1999; 99WO-US11814.
XX
XX 28-MAY-1998; 98US-0087080.
XX 28-AUG-1998; 98US-0141908.
XX 22-SEP-1998; 98US-0100880.
XX 08-FEB-1999; 99US-0119139.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Ashley G, Bellach MC, Bellach M, McDaniel R, Tang L;
XX
XX WPI: 2000-072618/06.
XX P-PSDB: AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
XX AAY67208, AAY67211.
XX
XX New recombinant DNA encoding a domain of narbonolide polyketide
XX synthase, for production of ketolide antibiotics -
XX
XX Example 2: Page 16-27; 98pp; English.
XX
XX This is the recombinant cosmid PKOS023-27 DNA sequence which contains a
XX Streptomyces venezuelae DNA insert. The cosmid contains open reading
XX frames which encode the various modules of the narbonolide polyketide
XX synthase (PKS). The invention relates to recombinant DNA containing a
XX coding sequence for a narbonolide PKS. Polyketides are compounds
XX synthesised from 2-carbon units through a series of condensations and
XX subsequent modifications. Modular PKSs are responsible for the production
XX of many antibiotics including picromycin. The narbonolide PKS consists of
XX a loading module, six extender modules, and two thioester domains. Four
XX proteins make up the narbonolide PKS (PICAI, PICAI1, PICAI11 and PICAIIV).
XX PICAI includes the loading module and extender modules 1 and 2, PICAI1
XX includes extender modules 3 and 4, PICAI11 includes extender module 5 and
XX PICAIIV includes extender module 6 and a type II thioesterase domain. The
XX second type II thioesterase domain is found on the PICB protein. The
XX nucleotide sequences encoding all of these proteins can be isolated in
XX recombinant form from the recombinant cosmid PKOS023-27. Narbonolide is
XX desosaminylated in S. venezuelae to yield narbomycin. The desosaminyl
XX transferase enzyme is required for this conversion, and the desosamine
XX biosynthetic genes are also found in cosmid PKOS023-27. The recombinant
XX DNA of the invention is used to express, in transformed cells,
XX narbonolide (or its derivatives) or other ketolides (particularly
XX hybrids), which may then be converted (e.g. by other enzymes
XX recombinantly expressed in the same hosts) to polyketide antibiotics or
XX their intermediates. The antibiotics are useful in human or veterinary
XX medicine.
XX
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 36.6%; Score 37; DB 21; Length 38506;
Best Local Similarity 60.4%; Pred. No. 0.33;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 1 CGGTGCAAAATGCCGATGATAGGGCCGAGTGTCCGACGCGGACCCGTCACAGT 60
DB 23143 CCGTGCACTGCGCCGACACTGGGGCGGAGGTCCACGCGGACGAGTCACGGGAAGT 23202

OY 61 GGGCAGCGGTTTCAGGCGCATGGCGCTGCCGCGCAGCAGCATATC 101
DB 23203 GGGACGCGCTCGCGGCGCTGGAGCAGCGGCGCATATC 23243

Db 23203 GGGACGCCCTGGCGGCTGGCGTGGAGACGCCGACATC 23243

```
RESULT 14
ID AAA14651
AC AAA14651;
XX AAA14651;
XX 08-AUG-2000 (first entry)
DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
XX FK-520: polyketide synthase; PKS; gene cluster; immunosuppressant;
KM Streptomyces hygroscopicus var. airomyceticus; immunophilin;
KM FK-506 binding protein; polyketide compound; transplant rejection;
KM graft-versus-host disease; uveitis; alopecia universalis;
KM autoimmune chronic active hepatitis; inflammatory bowel disease;
KM multiple sclerosis; primary biliary cirrhosis; scleroderma;
KM neurite outgrowth; nerve regrowth; Parkinson's disease;
KM Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KM peripheral neuropathy; ss.
XX
XX Streptomyces hygroscopicus.
XX
XX Key Location/Qualifiers
FH complement (412..1836)
FT /tag- a
FT /note- "fkDM gene"
FT complement (2020..3579)
FT /tag- b
FT /note- "fkBY gene"
FT 3969..4496
FT /tag- c
FT /note- "fkBR2 gene"
FT complement (4495..5488)
FT /tag- d
FT /note- "fkBR1 gene"
FT 5601..6818
FT /tag- e
FT /note- "fkBE gene"
FT 6808..8052
FT /tag- f
FT /note- "fkBF gene"
FT 8156..8824
FT /tag- g
FT /note- "fkBG gene"
FT complement (9122..5883)
FT /tag- h
FT /note- "fkBH gene"
FT complement (9894..10994)
FT /tag- i
FT /note- "fkBI gene"
FT complement (10987..11247)
FT /tag- j
FT /note- "fkBJ gene"
FT complement (11244..12092)
FT /tag- k
FT /note- "fkBK gene"
FT complement (12113..13150)
FT /tag- l
FT /note- "fkBL gene"
FT complement (13212..23988)
FT /tag- m
FT /note- "fkBC gene"
FT complement (13452..13662)
FT /tag- n
FT /note- "ACP6"
FT complement (13761..14394)
FT /tag- o
FT /note- "KR6"
FT complement (14517..15294)
FT /tag- p
FT misc_feature
```

```
FT /note- "ER6"
FT complement (15438..16587)
FT /tag- q
FT /note- "dehydratase domain (DH) 6"
FT complement (16587..17820)
FT /tag- r
FT /note- "acyltransferase domain (AT) 6"
FT complement (17820..19053)
FT /tag- s
FT /note- "KS6"
FT complement (19116..19326)
FT /tag- t
FT /note- "ACP5"
FT complement (19464..20097)
FT /tag- u
FT /note- "KR5"
FT complement (20241..21420)
FT /tag- v
FT /note- "DH5"
FT complement (21420..22653)
FT /tag- w
FT /note- "AT5"
FT complement (22653..23892)
FT /tag- x
FT /note- "KS5"
FT complement (23992..46573)
FT /tag- y
FT /note- "fkDB gene"
FT complement (24163..24373)
FT /tag- z
FT /note- "ACP4"
FT complement (24997..26146)
FT /tag- aa
FT /note- "DH4 (inactive)"
FT complement (26146..27430)
FT /tag- ab
FT /note- "AT4"
FT complement (27430..28684)
FT /tag- ac
FT /note- "KS4"
FT complement (28750..28960)
FT /tag- ad
FT /note- "ACP3"
FT complement (29092..29740)
FT /tag- ae
FT /note- "KR3"
FT complement (29869..31018)
FT /tag- af
FT /note- "DH3 (inactive)"
FT complement (31018..32185)
FT /tag- ag
FT /note- "AT3"
FT complement (32185..33439)
FT /tag- ah
FT /note- "KS3"
FT complement (33505..33715)
FT /tag- ai
FT /note- "ACP2"
FT complement (33823..34480)
FT /tag- aj
FT /note- "KR2"
FT complement (34606..35749)
FT /tag- ak
FT complement (28750..28960)
FT /tag- al
FT /note- "DH2 (inactive)"
FT complement (35749..37144)
FT /tag- am
FT /note- "AT2"
FT complement (37145..38296)
FT /tag- an
FT /note- "KS2"
FT complement (38371..38581)
FT misc_feature
```



```

FT      /*tag= ac
FT      /note= "ACP1"
FT      complement (38677..39307)
FT      /*tag= ap
FT      /note= "KR1"
FT      complement (39442..40609)
FT      /*tag= aq
FT      /note= "DH1"
FT      complement (40609..41842)
FT      /*tag= ar
FT      /note= "AT1"
FT      complement (41842..43093)
FT      /*tag= as
FT      /note= "KS of extender module 1 (KS1)"
FT      complement (43144..43660)
FT      /*tag= at
FT      /note= "ACP of loading domain"
FT      complement (43777..44629)
FT      /*tag= au
FT      /note= "ER of loading domain"
FT      complement (44974..46573)
FT      /*tag= av
FT      /note= "CoA ligase of loading domain"
FT      46754..47788
FT      /*tag= aw
FT      /note= "fkbp gene"
FT      47785..52272
FT      /*tag= ax
FT      /note= "fkbp gene"
FT      52275..71465
FT      /*tag= ay
FT      /note= "fkbp gene"
FT      52362..53576
FT      /*tag= az
FT      /note= "KS7"
FT      53577..54716
FT      /*tag= bb
FT      /note= "AT7"
FT      54717..55871
FT      /*tag= ba
FT      /note= "DH7"
FT      56019..56819
FT      /*tag= bc
FT      /note= "ER7"
FT      56943..57575
FT      /*tag= bd
FT      /note= "KR7"
FT      57710..57920
FT      /*tag= be
FT      /note= "ACP7"
FT      57990..59243
FT      /*tag= bf
FT      /note= "KS8"
FT      59244..60398
FT      /*tag= bg
FT      /note= "AT8"
FT      60399..61412
FT      /*tag= bh
FT      /note= "DHH (inactive)"

Query Match      36.6%; Score 37; DB 21; Length 77536;
Best Local Similarity 60.4%; Pred. No. 0.33; Indels 40; Gaps 0;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Oy      1 CGGTGCAATGCGCCGATAGGGGCGAGGTTGCGCCAGCGCGCCGTCAGT 60
Db      56473 CGGCACAGATGCGCCGACCTGAGCGCGAGCTCTAGCGCCAGTACCGGCAAGC 56532
Oy      61 GGGCAGCGTTCAGGCGCATGGCGCGCGCGCAGCAGCATC 101
Db      56533 AGCACGTCGTGCGCGCGCGGCGTCCGACAGCAGCATC 56573

```

```

RESULT 15
AAQ46806
ID      AAQ46806 standard; DNA; 29879 BP.
XX
AC      AAQ46806;
XX
DT      22-DEC-1993 (first entry)
XX
DE      eryA region of S. erythraea chromosome.
XX
KW      Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW      erythromycin; condensation; elongation; acyl chain growth;
KW      gene replacement; ss.
XX
OS      Saccharopolyspora erythraea.
XX
FH      Key
FH      CDS
FT      Location/Qualifiers
FT      3..10199
FT      /*tag= a
FT      /note= "ORF 1"
FT      10218..20921
FT      CDS
FT      /*tag= b
FT      /note= "ORF 2"
FT      20922..29879
FT      CDS
FT      /*tag= c
FT      /note= "ORF 3"
XX
XX      WO9313663-A.
XX
XX      22-JUL-1993.
XX
XX      17-JAN-1992; 92WO-US00427.
XX
XX      17-JAN-1992; 92WO-US00427.
XX
XX      (ABBO ) ABBOTT LAB.
XX
XX      Donadio S, Katz L, McAlpine JB;
XX
XX      WPI: 1993-242804/30.
XX      P-PSDB; AAR44430-32.
XX
XX      Biosynthesis of specific polyketide analogues esp. erythromycin
XX      cpds. - by introducing altered biosynthetic gene-contg. DNA into
XX      microorganisms
XX
XX      Claim 27; Fig 2; 133pp; English.
XX
XX      This sequence represents a fragment of the Saccharopolyspora erythraea
XX      genome, designated eryA. The polypeptides encoded by this region
XX      are involved in the biosynthesis of the polyketide segment of
XX      erythromycin. eryA is organised in modules and each module takes care
XX      of one condensation step. The precise succession of elongation steps
XX      is dictated by the genetic order of the modules. This fragment may be
XX      specifically altered such that novel polyketide molecules of desired
XX      structure are produced. Three types of alteration may be produced:
XX      CC structure inactivating a single function in a module which does not arrest
XX      acyl chain growth; those inactivating a single function in a module
XX      CC which does affect chain growth; and those affecting an entire module.
XX      The mutations may be introduced by gene replacement.
XX
XX      Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
SQ
Query Match      35.6%; Score 36; DB 14; Length 29879;
Best Local Similarity 62.0%; Pred. No. 0.6;
Matches 57; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Oy      10 TCGCCCATGATAGGGGCGAGTGTGCGCCAGCGCGCCGTCACAGTGCAGGG 69
Db      19147 TGGCCCGCGAGCGGGGCGGAGGTGTGCGCCAGCGCGCGCGCAAGCAGGGAGCG 19206
Oy      70 TTCAGGCCATGGGCGCGCGCGCAGCAGCATC 101

```

Wed Nov 6 09:54:50 2002

us-09-724-876-2_copy_29800_29900.rng

Page 14

Db 19207 TCGGGGCTCGGCTCGACGACGAGCAATC 19238

Search completed: November 5, 2002, 14:00:26
Job time : 1436.22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds

(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900

Perfect score: 101

Sequence: 1 cgtgtcaaatccgccgcatgtg.....gcgtgcgcgcagcacatc 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	68750	1 AF210843	AF210843 Sorangium
3	101	100.0	68750	6 AR193029	AR193029 Sequence
4	101	100.0	68750	6 AR199551	AR199551 Sequence
5	101	100.0	68750	6 AR199559	AR199559 Sequence
6	101	100.0	68750	6 AR199567	AR199567 Sequence
7	101	100.0	68750	6 AR201097	AR201097 Sequence
8	101	100.0	68750	6 AR208671	AR208671 Sequence
9	101	100.0	71989	6 AR172664	AR172664 Sequence
10	49.8	49.3	41097	1 AF016585	AF016585 Streptomy
11	49.8	49.3	123580	1 AF263912	AF263912 Streptomy
12	49.8	49.3	125401	6 AX211739	AX211739 Sequence
13	48.8	48.3	66808	1 SM0421825	SM0421825 Stigmatel
14	48.4	47.9	30000	6 AX250261	AX250261 Sequence
15	48.4	47.9	30000	6 AX250262	AX250262 Sequence
16	48.2	47.7	41782	1 AV118081	AV118081 Streptomy
17	46.6	46.1	65140	6 AX211705	AX211705 Sequence
18	43.4	43.0	3225	1 AF233752	AF233752 Mycobacte
19	41.8	41.4	104326	1 AB070940	AB070940 Streptomy
20	40.8	40.4	1784	1 AF41157353	AF411575 Actinomad
21	40.8	40.4	47981	1 AF263245	AF263245 Micromono
22	40.8	40.4	47981	6 AX112026	AX112026 Sequence
23	40.4	40.0	30000	6 AX250263	AX250263 Sequence
24	40.2	39.8	113193	1 AF357202	AF357202 Streptomy
25	39.2	38.8	15970	1 AE007061	AE007061 Mycobacte
26	39.2	38.8	53450	1 MTV018	MTV018 Mycobacte
27	38.8	38.4	32870	1 AF007101	AF007101 Streptomy
28	38.2	37.8	82746	1 AF453501	AF453501 Actinosyn
29	37.2	36.8	18614	1 AE007033	AE007033 Mycobacte
30	37.2	36.8	38000	1 MTG06H11	MTG06H11 Mycobacteri
31	37	36.6	37948	1 AF079138	AF079138 Streptomy
32	37	36.6	77534	1 AF235504	AF235504 Streptomy
33	36	35.6	20235	1 SERERYAB	SERERYAB M63677 S.erythraea
34	36	35.6	20235	6 AR049368	AR049368 Sequence
35	36	35.6	20235	6 AR095529	AR095529 Sequence
36	36	35.6	20444	1 SERERYABS	SERERYABS X62569 S.erythraea
37	35.6	35.2	6459	6 AX089458	AX089458 Sequence
38	35.6	35.2	16096	1 AE007124	AE007124 Mycobacte
39	35.6	35.2	24292	1 MTG0724G1	MTG0724G1 Mycobacteri
40	35.6	35.2	43280	1 SFU78289	SFU78289 Streptomyce
41	35.6	35.2	50000	6 AX089417	AX089417 Sequence
42	35.6	35.2	50000	6 AX089420	AX089420 Sequence
43	35.6	35.2	80161	1 AV007564	AV007564 Saccharop
44	35.6	35.2	80161	6 AR165018	AR165018 Sequence
45	35.4	35.0	14402	1 AF098795	AF098795 Pseudomon

ALIGNMENTS

```
RESULT 1
LOCUS AF217189
DEFINITION AF217189 58733 bp DNA Linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS Polyanthum cellulosum.
SOURCE Polyanthum cellulosum
ORGANISM Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyanthaceae; Polyanthum.
REFERENCE 1 (bases 1 to 58733)
```

AUTHORS	Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE	Cloning and heterologous expression of the epothilone gene cluster
JOURNAL	Science 287 (5453), 640-642 (2000)
MEDLINE	20115953
PUBMED	10649995
REFERENCE	2 (bases 1 to 5873)
AUTHORS	Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE	Isolation and characterization of the epohtilone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL	Gene 249 (1-2), 153-160 (2000)
MEDLINE	20293058
PUBMED	10631849
REFERENCE	3 (bases 1 to 5873)
AUTHORS	Julien,B.
TITLE	Direct Submission
JOURNAL	Submitted (16-Dec-1999) Kowan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES	Location/Qualifiers
SOURCE	1..58733

source
names
location/Qualifiers
1..587733
/organism="Polysiphium cellulosum"
/strain="Smp4"
db_xref="taxon:56"
1..992
/note="Orfa"
/codon_start=3
/transl_table=11
/product="putative transposase"
protein_id="AAF6:878.1"
/db_xref="gi:7453:55"
/translation="VRGHVEAFADFGVPRVLVLYDNLKNAYVERHGDAIRRHPLLAL
SADYFEERPAVAVARGNIKGRVERAIRVRGCFEERAYADIGOLINQAEWTSALAL
DSMWEDVARYRQAFDERSVLYLHPTPTPPDEHREVEVGHKTPVARELNDVSYH
DSTRRTVLADISQVRIADGNOIATVIRSMDOGQIQEENHQRIVDEKRRAREHR
HGIDRAARRASQAFRLRIVERGDWGSALIRLLQDLDAVAAALEALVELEDRDTI
GIAVRQYARSSRERHLPVPSIVPTGEEHAALVTVPHSLTFYDALKKDPTP"
989..1501
/note="Orfg"
/codon_start=1
/transl_table=11
/product="putative transposase"
protein_id="AAF6:379.1"
/db_xref="gi:7453:56"
/translation="MTDLPTPETKDRKSLGIFGLACMQDLADKPMLELYALIEERE
HKRSLEERRKSRVAAECPMTDFSSWPKIDREAVYDLDLSRADLFEVYVRD
AOKPLLTSTNKKFADWQGFPPHACVYTVLDRVLHRAVEVIEEASRYRIKAKELNAT
RTKORRTKKH"
1998..6263
/gene="epoa"
1998..6263
/note="type I polyketide synthase; epothonone PKS loading
module"
/codon_start=1
/transl_table=11
/product="epoa"
protein_id="AAF6:380.1"
/db_xref="gi:7453:57"
/translation="MADRITERRADDPRIATYAGACRLLPGVYDLSGFWTLLEGSROY
GOVPERRDAAWPPDPLAGKPTPTASTLSLDVACDEADPSAFSEIRALRMDPAIR
DLERCHALEENHATPSFLVGTETGVGTGSPSTACBLAPRAISHEIDHIGDLGT
PSVGRGRTSYVLGRCRCAYVDATSSSLVAHLCCSLRGESCTIALAGVSSIMLSP
SLTVLSTRALADGRCRHAFSAEDGSGREGGAVYLRKLSARADGDTLIVINGE
SAINHDGASGTLVPGNSQELVLRKALADGCASSGVYEAGTGTGLDLEIQA
LNAVYGLDRADVTLIGSVKTNLHGFSYAGTIGLLKLVVLSLOHGIIPALHQAOLN
PIISMGIDRLVTRTARPIPMPMNTPRRAGVSSFGMSTNAHVLEEAAPATCTPAPPE
PRAELTVLSARTAAALDHAARLRDLHETVPSQCLGDVAESLATTBSAMEHRLRAV
SSEGLRAALDAAAGOTPPGVYRGIASSRCKLAFLEFGQAGOTLGMGRGLYDVAPE
REAPCLRRLNQEOLDRPLREYMAEPASVDAALDQTAFTQPALFTFEYALALAMS
GWGEELVAHSGSITGLVAVYAGVLEDAFLVAAQRGLMQALPACGAMVSIAPBA
DYAAALVVAHGSIVAIYANJPDQVVIAGAGCPVHAIAAAMARGATKALVSHAFHS
VYALHMAAGFCGVEVSEVSYRRPSITLVNLSGKAGTDEVSFGVVRHAREVYFAAGC
VYALHMAAGFCGVEVSEVSYRRPSITLVNLSGKAGTDEVSFGVVRHAREVYFAAGC

gene
CDS

ANGCIVSMASRDLPGSGRRVLPPLPYPMQREWYITDRKADDAARAGDGRAGACDDEVEG
GAVRCDGRNSRLDHPPEGRGEEKYEAAGDPRFEITDEPVDLRTATERRAG
LGEVETINDANGKLEFNDVQALCMVPDDLPGRKNPPLLILGEGCAGRIIVAVEGYNGLY
VQCPILIASACAEPIHTVTSALPLPQALSTTEKAAEVATLTMALSDILRLDLP
GRTVLIHATAGCVLAVQMHGAEVAHATATPEKRYLESILVRSVRSNDRV
ADVRAATGECGEMDVNLSTLGEGLIDSFNILSRHGFEVLEGRDCAVQNOGLFPLP
NLSEFLVDLRCMMEERARVAREBELLSIAGVETPPIATPLIAVAAQAFSMQ
AOLKRLVETLIGDEEVOIRIPTNAGSGPSTGMDLDRASAPAPARAALAEALTRQ
VSQHTPEITVEIGVAEAFTRIGMDSLMAVELNRILAEASLKLSTTFSTSPNTALLTQ
QULDLATLLEBRVAENLRAGVQDFVSSGQDDWEIIAL"
6260..10492
/gene="gpoB"
6360..10492
/gene="gpoB"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"

[illegible]

gene		OTSTRMWTTLDLKRLPMLGDHROVAVPEGAAYLEMAISSGAEALDGPLOITDVA LAEALFAGDAAYVQVYVTTTQPSGRLOFIASPRAGASHFEVHARGALLRERE VPAGITLSAVARLOASMPAAATYAELTEKMOIOPAGPOGIAELTEKMOIOPAG DAAGSAEYRLHPLALDACFOVVSILGAGGEMPMVAVSGSRILOQSGEIMCA RVVHGKQTPDROGADFWYDSSGAVVAEYSGVAOQLPGCVRREREDDFLEWEP AAVGTAKVNGRWLLLGSGGGLGALRSMLEAGHAAVHAESNTSAAGVALLAKAF DQAPPTAVHGLSGIDGGELDPGAGAGADSVTOQALDGLRVIAMEHADICARVDL QALAGMFRDAPRLMLLTGAGAGAGADSVTOQALDGLRVIAMEHADICARVDL PTREDGLGALLAEILDAAEAVALGAGRCVARIYRROPETPRGCRISCVPTDVT IRADSTYLLGGJGGLSVAGNLAEEGACILVYVGSAGVAYQRAVVALEHGRK VYAKADVADRAQLERLREVTTSGMFLRVHAGLIDGLMOQTPARKVMARK VOGALHLHALTREAPLSFVLYASGVGLSPGQGNAAANTFIDLALHRRAGGLA LSYDMGLFAEAGMAAOEDRGARLVSGMRSLLPDEGLSALRLIESGRVQVGMPLN PRLAVELYPAASRMLSLVTAHARASGAGDGLDLRLAAEPGARSGLLEPLR AOISOVRLPEGRKEVDAPLSIGMNSLMGLRLNRLEAMIGITVPATLLMTYPTVA LSGLHAREAEAPAVESPHTTASAVTEEMSQDDTLQILAAKRALT"
CDS		16134..37907 /gene="epod" /protein_id="AAE26904.1" /note="type 1 polyketide synthase; epothilone PKS modules 3, 4, 5, and 6" /codon_start=1 /transl_table=11 /product="epod" /protein_id="AAE26904.1" /db_xref="GI:7453560"
CDS		/translation="MTTRGPTAQQNPDKQAIIITQLEERLACLAQALEPTEPIAY GICRREPGADAPAEWELLDAERDAVQPLDRRALVGAPEVAEVPMAWILLTEPIDC FDAEFGISPREARSLDPORHLLEVMELIDGIPRSIDGSRTPVFGAFADPA RTVATRLPREEDAVSATGNMISIAAGRLSYTLGLOGCLTVDPACSSILVAHLACS LRAGEEDLALAGVSTLSPDMEAAARTOALSCRTGEDASANGFVEGEGGLVY LKRLSDAQRDGRIMALIRGSAINHDSRGITRAPVLAQETVIREALREARVAGAV DYVTHGTGTSIDPREVEYERATVGPARSDDRCVCYAKTNGHLEAAAGVAGLIK AALSLTHERIPRLNFRITLNLRLKLEGSALALATEVPWPRTDPRFAGVSSFCMSGT NAHVLEAEAPAVELEMPAPERSSELVLSEKSEGLALSAVAGOTPAQAARCIASSSGKLALFT AFSLATRSAMTHRLAAVATRSBGLLALSAVAGOTPAQAARCIASSSGKLALFT GGGAQTPGMGRGLCAAMPAREAREDCVTELDREILREVMAEAGSABSLIDDT AFQOPALFAVEYALTLALMRNGVEPELLVGHISGEIYAAVCYFSLDECVRYAAG RLMGLSAGGAMVSIGPEAEVAANAAPHAAVSIANAANPBOVYVINGVEDAVATIA GFARGVTRKRLHSHAFSHSLPEMLPEEGRVAASVYTRRPSVLSVNSLKGVTDE
Query Match		100.0%; Score 101; DB 1; Length 58733; Best local similarity 100.0%; Pred. No. 1,6e-13; Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	1	CGGTGCAAAATGCGCCGATGATAGGGCGGAGGTCTGCCACGCGGAGCCCGTCAAGT 60
Db	29800	CGGTGCAAAATGCGCCGATGATAGGGCGGAGGTCTGCCACGCGGAGCCCGTCAAGT 29859
Oy	61	GGGACGCGGTTCAGGCCATGGCGCGCGCGGCGGACGACATC 101
Db	29860	GGGACGCGGTTCAGGCCATGGCGCGCGCGGCGGACGACATC 29900
RESULT 2		AF210843
LOCUS		AF210843
DEFINITION		Sorangium cellulosum strain so ce90 epothilone biosynthesis gene cluster, complete sequence.
ACCESSION		AF210843
VERSION		AF210843.1
KEYWORDS		AF210843.1 GI:6724237
SOURCE		Polyangium cellulosum. Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangiineae; Polyangiaceae; Polyangium.
ORGANISM		Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milamov, M., Cyrr, D. D., Gottschalk, J., May, J. M., Hu, A., Goff, S., Schmidt, J., and Ligon, J. M. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum so ce90
REFERENCE		
AUTHORS		
TITLE		
CDS		

```
/product-"unknown"  
/protein_id-"AAF6907.1"  
/db_xref="GI:6724241"  
/translation-"MDHPIGRRCALAEQYIPGSGNGPEPOMTSHEATCLNABSDA  
OAVATVYFSDRDPAGPYRVTVPARTRRHVFNDLTEPEIPIDDTYASVIESDAPIV  
OHTRLDSRQAEALLSTIAYTDRE"  
6226..6675  
/note="ORF5"  
/codon_start=1  
/transl_table=11  
/product-"unknown"  
/protein_id-"AAF26908.1"  
/db_xref="GI:6724242"  
/translation-"MKHVDTCGRGRRGRIHTGLGLASMLACGSPSEKTYOQTRLAP  
GADARVTAVDPAATTTALAVDVVHLSPERLEAGSERVVMORPEPSPMRVGVLD  
VYADSRGKLAETVYVYANFELLITAEKQSSPSPSAVIGTISVG"  
7610..11875  
/gene="epoa"  
/gene="epoa"  
7610..11875  
/gene="epoa"  
/note="EPoS A"  
/codon_start=1  
/transl_table=11  
/product-"polyketide synthase"  
/protein_id-"AAF26919.1"  
/db_xref="GI:6724253"  
/translation-"MADPIETRAEDPIATVASCRLPGCVTDLSGFWTLGSRDVT  
GVPAERMDAAAFDDPIDPGRTPYTRASFISDVACFPASFFGISPREALMDPPIHR  
LLEVCWEALNAALAP:ALVGTETGVIGTIPSEYBALPQATSAEIDAHGGLGTM  
PSVAGRIYALGLRQCAVADPTAVSSSLVAHLAQSLRSGECSTALGAGVSLITSP  
STLWMLSKTRALARDCRKAFASEADGFEGRGCAVYLKRGADGDRLLAVIAG  
SAIINDGASSGLTPNGSQEILKALDACCAASVGVYBAHGCTTLLGDTIEQA  
LNAVYGLRDVATPPLIGSKTNLGPVFAASITGLIKVLSLQHQIPAHLLAALN  
PRIWGDRLVTRATWPDWNTPRACVSSFGMSNAHVLEAPATCTPPAPE  
RPAELLVLSARNAALDQAARLRDLLETYPQCLGDVAFSLATTRSAENHRLAAAT  
SREGRALAAAGAGTSPGAVRSIADSSRGKLAFLTQGAQTLQMGGLDVSAF  
REAFDLCVRLFNQELDRFLREVMMAEPASVDALLDQTAFTQPALTEFEVLSIAALRS  
WCEPELVAGHSIGELVACVAVFSEDAVFLVARGMLPAGMGAVSIEAPEA  
DVAAVAVPHAAVSIAAVVAPDQVVIAGQSCVHAIAAMARGAATKALHVSHEFS  
PLMAPLEAFGRVAVESVSRPSIVLVNSLQKACIDEVSSPGYWRHRAHREVRADG  
VKALHAAGAGTVEVGPKSTLLGLVPACMDARPALASSRAGDEPAPVHALGLIM  
AVGVLVSMAGLFPSSGGR/PLPTYPQORRYYIDTRKADDAAGDRAPACGHEVEEG  
CAVRGDRRSARLDHPP33GRREKKEAGDPFLEIDPEVLDHLVYTERERAPG  
LCEVEIAVDAALSFENDVLLGAWDDLPCKNPDLILGEGCAGIVAGGVNGLV  
VGPVIALSAGAFVHTVSALVLPKPOLSAIEAAMPVYLYRWYLDRIARLOP  
GRRVLIHAATGCVGLAAVMAOHVGAEVATGTPKRAVLSLIGRVYSDRSDFP  
ADVRAMTGEGVDVYVNSL/SGELIDKSFLLRSHGFEVLGRDCYADQQLRPFIL  
NLFSFIVDRGMLEBRPAIVRALLEELIGLIAAGFTTPTPIATPIARVAALEAFRTQ  
AOLHCKIVTIGDPEYQIILPTPHAGAGPSTGDRDILDRASAPARAALAEAFRTQ  
VSQVLRTPETIKYGAELP/RLGMDSIMAVELENRITFASIKTKLSTTFLSTSNILLA  
ONLLDAIATLSLEVAANLNLGAVQNDVSSGADQDREITIL"  
7640..8920  
/gene="epoa"  
/note="Region: beta-ketoacyl synthase/malonyl-CoA  
decarboxylase"  
9236..10201  
/gene="epoa"  
/note="Region: acetate acyltransferase"  
10529..11428  
/gene="epoa"  
/note="Region: enoyl reductase"  
11549..11764  
/gene="epoa"  
/note="Region: acyl carrier protein"  
11872..16104  
/gene="epop"  
11872..16104  
/gene="epop"  
/note="EPoS P: confuins condensation/heterocyclization,  
cysteine adenylation, putative NAD(P)H oxidase, and  
peptidyl carrier domains"  
/codon_start=1  
/transl_table=11
```

```
/product-"nonribosomal peptide synthetase"  
/protein_id-"AAF26925.1"  
/db_xref="GI:6724259"  
/translation-"WTINQLNELEHGKIKLAADGERLQIAPKNALNPMLLARISEH  
KSTITLMLRQRPASIEIVPAPAEHRAHPPLTDIDESYVLGRGATVPSGIHAAREYD  
CTDLDPRLSRARFVARHDMLRHATPLDMKQVLEPVDIIDLIDRGLDRSREA  
RLVSLDAMSRHRIYDTERPPLVAVVAALDEKQVRLVSLIDINVLDSISITKDWL  
SFYEDPETSLPVLEIYRNDYVLRLESRKSEHNSMDYKRIELNEPPTLPKAD  
PSTLEIKFRFRTQEWLPDSWGLRALKRSGEHLPTGTIILAFSEVIGRMSAPFTL  
NITLFNRLVPHRNVNDITGDEFTSMALLIDITRTDSEFORAKRIODEQLMEAMDHDS  
GIEVOREARAVYGIORGALFVVLVLSALNOQVAVTSIQRIGTVPYSTQTPQLLDH  
QLYEHGDVILAMDIVDVPFDDLDMDLLEAVVFLRKTTEPMGEQVRCGLPQALE  
EAGAPNPTLVAAVMEKGMQVAVVAVLESAAVYPIADAPAEIRIYVLTSGTGKLV  
LTQPMLDGKLSWPPGILRVLSEAGVEGDGQPPMPPIOTPSDLAVIYVTSGLPK  
GVMIHRGAVNTIIDLINREFEIGPDRLVLALESFSDLSYDVDFGILAAAGTIVVPA  
SKLRDPAHMAELIEREKVTVMNSVPALRMLVEHEGRSPDSLARSRLSLSGWIPV  
GLPGELOAIRPGVSVISLGATSEASITMSIGYPVRVDSLMASTIPYGRPLRNOTFVLD  
BALERPVPWPGQLYIGGVLALGWRBDEKTRKSELVHPTEGRLYKTSGLGRYLPD  
GNIFEMGREDDNIGIKRGYRVELGEIEETLKSHPNVNDVIVYVGDANDANKLLAYVVD  
EGRRRARABODASLKTERRIDARAHAAEADGSLDGRVOFKIARHGLRDLGKPYVD  
TGODPREAGLDVYARRSVRFTLEAP/PEVEGRSLCTLSVPEPGATLPKPRYPSAG  
STYPVOYAYVYSGRIIPGVDEGFTYYHPEPFRILKLSHGTERCAHVNQNDVDEEA  
PULLIYGRIDAEISIXGSSSREPCILGAYNOLIMROAPSCNTGVCYCGPNEVOR  
PYLDIRHSDVYHGMCLGRNDRPQDQCTICODSSPRASTRGARGREQFAMLRD  
FLRKLPEIMVPTVVEEDALPLITNGVYDKALDERDITSPRHSQTRARDALEI  
LVAVYREVIGLEVDGLQSFYDLGATSHIYAKRSLDOKRLDREIATITELQYPLGS  
LASGLRSDSLDQRFNMQDRVRRKGRRS"  
16251..21749  
/gene="epob"  
16251..21749  
/gene="epob"  
/note="EPoS B"  
/codon_start=1  
/transl_table=11  
/product-"polyketide synthase"  
/protein_id-"AAF26920.1"  
/db_xref="GI:6724254"
```

```
Query Match 100.0%; Score 101; DB 1; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 1,56-13;  
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGTGCAATCGCCCGCATGAGGCGCGAGGTGTCGCCAGCGGACCCGTCACACT 60  
Db 35412 CGGTGCAATCGCCCGCATGAGGCGCGAGGTGTCGCCAGCGGACCCGTCACACT 35471  
QY 61 GGGCAGCGGTTGCGGCATGGCGGTGCGCGGACGACATC 101  
Db 35472 GGGCAGCGGTTGCGGCATGGCGGTGCGCGGACGACATC 35512
```

```
RESULT 3  
LOCUS ARI93029 68750 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6346404.  
ACCESSION ARI93029  
VERSION ARI93029.1 GI:20238994  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)  
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and  
Goriach,J.  
TITLE Genes for the biosynthesis of epoethiones  
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;  
FEATURES Location/Qualifiers  
source 1..68750  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN
```

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 60
Db 35412 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 35471

Qy 61 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 101
Db 35472 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 35512

RESULT 4
AR199551 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199551
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 1159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 60
Db 35412 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 35471

Qy 61 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 101
Db 35472 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 35512

RESULT 5
AR199559 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199559
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 1159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 60
Db 35412 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 35471

Qy 61 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 101
Db 35472 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 35512

Db 35412 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 35471

Qy 61 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 101
Db 35472 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 35512

RESULT 6
AR199567 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199567
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 1159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 60
Db 35412 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 35471

Qy 61 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 101
Db 35472 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 35512

RESULT 7
AR201097 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR201097
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source location/Qualifiers
1. .68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 1159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 60
Db 35412 CGGTGCAAAATCGCCCATGTGATAGGGCGCGAGGTGTTCCGACGGCGAGCCCGTCCAAGT 35471

Qy 61 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 101
Db 35472 GGGCAGCGGTTCAAGGCGCATGGCGTGCCTGGCGGCGACGCACATC 35512

Db 35472 GGGCAGCGGTTGAGCCATGGCGTCCG:3GCACGCACATC 35512

RESULT 8
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE unknown.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnepf, T., Ligon, J., Madison, J., Molnar, T., Zitzke, R., Cyr, D., Dawn, and Goriach, J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1,5e-13;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGAATGCGCCGATGATAGGGCCGAGGTGTGCGCAGCGGAGCCGTCACAGT 60
Db 35412 CGGTGAATGCGCCGATGATAGGGCCGAGGTGTGCGCAGCGGAGCCGTCACAGT 35471

Qy 61 GGGCAGCGGTTGAGCCATGGCGTCCG:3GCACGCACATC 101
Db 35472 GGGCAGCGGTTGAGCCATGGCGTCCG:3GCACGCACATC 35512

RESULT 9
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE unknown.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien, B., Katz, L., Khosla, C. and Tang, L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1,5e-13;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGTGAATGCGCCGATGATAGGGCCGAGGTGTGCGCAGCGGAGCCGTCACAGT 60
Db 29800 CGGTGAATGCGCCGATGATAGGGCCGAGGTGTGCGCAGCGGAGCCGTCACAGT 29859

Qy 61 GGGCAGCGGTTGAGCCATGGCGTCCG:3GCACGCACATC 101
Db 29860 GGGCAGCGGTTGAGCCATGGCGTCCG:3GCACGCACATC 29900

RESULT 10
LOCUS AF016585 41097 bp DNA linear BCT 07-DEC-1997

DEFINITION Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nid1) gene, partial cds; polyketide synthase modules 1 through 7 (nidA) genes, complete cds; and N-methyltransferase homolog gene, partial cds.

ACCESSION AF016585
VERSION AF016585.1 GI:2558836
KEYWORDS
SOURCE Streptomyces caelestis.
ORGANISM Streptomyces caelestis.
REFERENCE 1 (bases 1 to 41097)
AUTHORS Kakavas, S., Katz, L., and Stasyl, D.
TITLE Identification and characterization of the nidamycin polyketide synthase genes from Streptomyces caelestis
JOURNAL J. Bacteriol. 179 (23), 7515-7522 (1997)
MEDLINE 9393718
PUBMED 98053867
REFERENCE 2 (bases 1 to 41097)
AUTHORS Kakavas, S. and Stasyl, D.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1997) 47P, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064, USA

FEATURES
source 1..41097
/organism="Streptomyces caelestis"
/strain="NRRL-2821"
/db_xref="taxon:36816"
/complement(<1..396)
/gene="nid1"
/complement(<1..396)
/gene="nid1"
/note="orf6; similar to tyl1"
/codon_start=1
/product="cytochrome P-450 hydroxylase homolog"
/transl_table=1
/protein_id="AAC46023.1"
/db_xref="GI:2558837"
/translation="MVDSVTGPEMELSDANAKELIDFNSNRTHHPVWDEGQAMOV
FRDYDYLTVSNHPEFSSDFEVAPEPELEMLGPGTIGALDPAHGPRIKLVSAQF
TPRMAQEQORIRVIAELLDVRGOKTA"
627..13649
/gene="nidA1"
627..13649
/gene="nidA1"
/note="orf1"
/codon_start=1
/transl_table=1
/product="polyketide synthase modules 1 and 2"
/protein_id="AAC46024.1"
/db_xref="GI:2558838"
/translation="MAGHGDTATKADDAEKSESDPAIAVIGSCRFPGAPGAIEFW
QLSSGADAVYTTADGRRGTIDAPAFEDAAFGMSPREAATDPOORLVEIGMEYL
EDAGTIPESLRGEASVFGVGMNDVYTLIHRGAFPDVTATGLOHSMANRLSYEL
GIRGASLVVDGSSSLVAVAVESLRGTSGLAGLVAVLVAERGSAMNRYGAL
SPDRCHTFDARANGYVRGGATLVVLRPLADALADDRVYCVVARGVATNDGGPGL
TVPRRGAQENTIRACDQACVPRADVAVRFLVGLGCTGAGDPVRAEALGAYGGRPN
EPLVGSVKNTIGLIRGAGIAGVRAALCHRPAPASINFEPPAPATIPERPAKY
OTANALORGGGPRILAGVASRGMGTNHYLETPGGRORAPENACIPSPAKY
MILSARSQALRAQARRRHELDSDADLDIATSLATRTREHRAVACFPDPRL
SSALDLAAGOTPRGVRIGSTDADGRALALFTGGAGAPGGOBLITTDHFAADE
VCEBLRGCTONLEVMFTPPQPLDRTEITOPALDALTALRTITATGDOAHVH
GSHVGEITTAHIAIGVLDLPDARLITRAHVMGQLPHGMALSVQAAHEDLDOLAHV
GVELTAANGPPLHCLSPRLALETAQLLEBONRHWMLVSHAFSALMDPMAGAR
DTLTAANGPPTLPILISNLGQIALDHPMLCTPDVMDHARVAFADVAHAGCTT
TYLEIGHPTLTLIHLTLDNPTTIPILARERPEPTLTGAIAGVATDGIIMAVIC
GASPRVRELPTVAFORRTWAPGLIPNAPADRPAPAEPRAAVGVSTSEALVRLVG
ETTASVGLGDPDEVALDRPFTSGLDSTMAVELAGLCTAGVATDPTLYVELPTPR
AVADHLAKTLIGESADADQEVNKRTEAEAKAGDPLAVIGICRFGGVATPDDILME
LVASGTDALSTFTPDGMDLDGLYDPPSPSTGSKSYVHGGLFDAAQFDFEFGISPR
EATMAGDQORLLETSEWALERAGVAVESLRGTSGLAGLVAVLVAERGSAMNRYGAL
GFLTGTASVAGSRIAYALGTRGPATVTDGATSSSLVLAHLVQSLRRECCDLALAG
GTVMSGPMFVEFSRORGLAPDGRKAFADADGTMAEGVGLTVERLSDAERLCH

VERSION AF263912.1 GI:8050835
KEYWORDS Streptomyces noursei.
SOURCE Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
ORGANISM Streptomyces noursei.
REFERENCE 1 (bases 1 to 123580)
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strim,A.R.,
Valla,S. and Zotchev,S.B.
TITLE Biosynthesis of the polyene antifungal antibiotic nystatin in
Streptomyces noursei ATCC 11455: analysis of the gene cluster and
deduction of the biosynthetic pathway
JOURNAL Chem. Biol. 7 (6), 395-403 (2000)
MEDLINE 20334850
PUBMED 10873841
REFERENCES 2 (bases 1 to 123580)
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
Valla,S. and Zotchev,S.B.
TITLE Direct Submission
JOURNAL Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
N-7489, Norway
FEATURES
source
1. 123580
Location/Qualifiers:
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
complement(46..784)
/gene="nysF"
complement(46..784)
/function="presumably post-translationally modifies ACP
domains on PKS"
/note="putative 4'-phosphopantetheine transferase"
/codon_start=1
/transl_table=11
/product="NysF"
/protein_id="AA07162.1"
/db_xref="GI:8050836"
/translation="MET..LPATVATEAAYDRPRPGRDLSSEREVITARAVERORE
FTVVRHARLRRLRGHPHRIALLPNRRGAPMPGIVSGMTGACGRAAASPAELISA
AVSIDAEPHPLPAGVLN.IALPSEPHYLVAALAHPRDVHMDLFLSAESVFKAMP
LVQRELDFSEAEIVDPTGAFTRALVPGPLGGRRVTFPGRMHSTPALITAVHL
PAPTPRDRERHTHLVN::PLPRPTFG"
complement(867..2484)
/gene="nysG"
complement(867..2484)
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysG"
/protein_id="AA07163.1"
/db_xref="GI:8050837"
/translation="MASPLDLEERTAPPARRLVGLLPHRRSVLAASMGVGLV
NMFGLLRGRTDLADYGLGPGAPGIDPAIGRLIVLAILVVASLMLAQR
LYASAVMRTIHELRLDARHKLRLPLRHDRDPAIGELISRTNDINDNCOOTLAE
LITSEILTMVLMLVLSPLAAVWLISVPASLIANISKRADPHYAAOAGSGL
NAHVEVCTGHALIKGFDRRAAEEREDCANVYRAAKKQASGAMPVMPVANT
GVYAVVAGMVKINGTITLIGVOAFITLAROESQIYELIASVAGRLSGVAVV
TLLDAPEOAPDLPRGTPARAGRVETDPSRYSPPDPLINISLTVRPGSTAVV
PTGACKRTGCLNLMPEYHDSRILLDCDPTATMTRDLSRFGILYDTHYEGGTIA
ENYAGARGAGCADIEMARATCADRFITLDCGIDVLDSDGSYSAGEROLIVAR
ATLARPAVIVIDEATSSVYRTVEVLQORMNSLRGRISFYIAHRSLITRDADLIVM
DGRIVGSGTHQQLCAQGLIARLHAARHTPTAGAAG"
complement(2662..4416)
/gene="nysH"
complement(2662..4416)
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1

/transl_table=11
/product="NysH"
/protein_id="AA071764.1"
/db_xref="GI:8050838"
/translation="MLRLRLRQLRPYMATRALVALDVLQTLITLPTGALLID
GVVRDGRITELGVVGVVALVQIAALGAALATATYAGRLRSALFRITIDS
AIEKGFSPILTRSVNDVQONIAQTRGELVVCAPLMCLGSVLLALRDVPAL
LVALVAVAVCGLLIARLGTLYARQTLIDGLRLLEATIGVAVNSFADHERA
RPAQNDPAFLVYSRRVRLIATMLPVLQLLMGFLVALLMDSHRIDAGRMFISLA
LTSLSLITMSVYMLAFELSVPRARVAGRAELAVLDGSSVAPPAAPQVPAGVARI
ELCAAGVRLTAAATLAAGVAVPQRPYLFSGTVAASLNRFRPDATDEELMALRVQA
RVGGVDVRELTAATLAAGVAVPQRPYLFSGTVAASLNRFRPDATDEELMALRVQA
ADFAVARMGDLDAETTOGGVNSGGORQLSLARALLRRLPEIYLFDDCFSLDQATDA
ALRTALVPTAGATVITVAORISAGRDADRIIVLDGRHVAVOGTSHVLLRTSPYREI
ALSQLEEEAAGLAGRP"
4714..5748
/gene="nysDIII"
4714..5748
/function="presumably involved in mycosamine biosynthesis"
/note="putative dGDP-mannose 4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NysDIII"
/protein_id="AA071765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITGONGSYIAEHLISQGYQWGLIRGOANPRSRV
SRISAEIDPDGDLMDGSIYAVDTPQDPPEYVINGATSPVMSQDAEITVEVQMG
VRLMEFATPMVSGLSSTRSPGRCQIRPEYQASSEMRCRAETPQRETTIRHPSPG
AAKAGHYTTRVRRPESRCMYAVSCMLPHEBSRGRGQEVYKXISLAVARIGQDRL
ALGNDAVRMGYACDVIYRAHMLIQDQAGDDYVYGTQMSVSRDAVIALRHCGLN
EDVYVIDPLVPRAEVYLCADSAQDRLGKPDVPTLLRMVVDLQVSRENO
YCDVLLAAM
5930..34363
/gene="nysI"
5930..34363
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="NysI"
/protein_id="AA071766.1"
/db_xref="GI:8050840"
/translation="MDNEOKLRDLKATADLRTRRHVKLSNAQEPVAILGTCR
YPGVRSPEDLRMVAVGEHGVTPPTDRGDLMLAAPTASGFLADADPDAF
GISPREAVAMPQOQVLESAEAEFAPGIDPTSVKSGRTGVIGAMADYRVPADG
ABGVLITGNTGSLVSGRTSYTPGVAVPVDVATDSSSLVAVHLATQILRAGECTIAL
AGGVITMGPGTFEIMGROGLSADGRCSFGDPTADGTGMAEGVITLVELSDAVRN
GHEILAVRGTFVNOGASNGLTAFNGSQOQVIOQALVNAARLAAGVIDVEAGTGT
TLGDEVEAOALATYAGNRPADRPILLTSVSNISNLSVAGVAVIKVMAMRGTL
PPTLAAEPTTHVDVSGCAVRLITDTTQMPATGAPRAAVSFGISGTNATYIEOAP
EPQEDAAATAODDAGSGPATPAPVAVGVVPLSGRPDALRGQAAALRAALDGRAP
DLIDLHSLATTBAGEFHRVAILATDHPALIDGJLALDADDPAAAPWITGTTRAE
RLAVLFTGQAGORGLAGRELAARPAATDALDADFTPLDRLPRLVINGTAAIL
DRTVAQAPLFAVEVALYRLIESFVRPDHLAGHSGVIAVANGSAVYLSCAENATVA
LAEOILASDGRTRRLRVSHAPSPMEMLDPAARVVDLLOPLIPVSNILNGKPA
TYAOLTAAGKRYDVRVNAVVRPFDGIDTAAHDTTARLITGDDVYLSMAOQDAADA
DAVTLDALPAPRBEHNTLTALAGLHVAISADMGSTLRLSISRTQHPIMADTRVC
PKATQSGTADLRSGILAAHPLISAAVYASLDAGSTLRLSISRTQHPIMADTRVC
TTLILGTPALELAVYAGDNCGRREETLAPLILPQGVQVQVQIAGNDVSGRR
VAVHARPDGDTPTATATGVLITTNDSRLPNSBEGGIFLAGDPPHRAADAKMP
AGAPLPDGGTDRILDADGVLITTNDSRLPNSBEGGIVAEVLEPEGRSDAEFGI
HAPLIDALAHAPFTGLDEGRGGLEFSWESVSLHAGATTLRLRITPADAALATY
ADTGAPVLSVDSILRSVATQOUDTAAVADALFRIDMTVPVPTATDPAALATY
DPFGILTTHAGPADAPATPYDIALAADSPVPTVYLDAGTDDAPARSAAHCAAE
ALAAVQTLUDHHERPAARLVFVTRGATGVADVAAAAGVLRSAQDSNPCCFALVLD
DPDAGVGAALVAAVLVSGEPOLAVRGLVRAVRLPRLTEVGAGADGTGGVADGSG
VSFSGEAGVLVYGTGAGVLAHLVAEYVRDILLVSRSGEAVYAGELVLAELAGV
GARVAVACDVYDRAAVVELVAGHVAVSVAHVAAGVLDGMAIGALTGELISVLRPKYD
AVMHLHEATRGIDIDAFVYFSSLAGVFSQCAQNAANAFLDALMTERRAEGLGILS

LAMEPQOSOMQIMTIDVAEELASAGVPLTSMOGLADPAAYMGDCAVCYRPL
 LPIVRAEGEYPPMLRSLIRVRAARAAVASSATRNGLADPAAYMGDCAVCYRPL
 OVALYLCHATGDDGDKARADRLDGFSLTAVELRNRLNVTGLRLPATLVDPYVHR
 LATYVLDLLEGLAEVAVYQOAPVAANADPYITVMCAQYPCQVSPEDMLRVLTEGT
 DAVSPGFLNMGVSELYHPDDPHSTYSTRSGGLHEGEDEPEFEMSERLALADY
 NOOKLLESMELIERGADYDPSLRSGPRVAGWYSPYSLMLASPEEFQSGSSSS
 PLASGAVATYLERGADYDPSLRSGPRVAGWYSPYSLMLASPEEFQSGSSSS
 AVYDPRAROKLSPDGRCKAFADADAGVSMSCGVYLERSDRNRNGHGLTAVNGR
 LAANODASNGLTAAPNPSOQYRTRALASGLYAGVDYVAHOTGQVTRPLTLEADQ
 LAYVGDREPERPLLTGKSNYTGQASAGVACIKVYLHAGVGRPLTRHVDASPS
 SHVMSGCAVELLEOAMRGESRVRASVSGSGFNHNYLYOAGACALAAAGAA
 ARRRGAPVILLSRGASALRGQANGLIGHCARDAEENHVALTATREFEORAA
 VVADROBLIASLGLAADRPDAVVEGSAKRGRTVLTGQSORAMERLEHETQ
 PERPAATDCAVCATPDLDPLEPREVEFADSDSLDETGMQPALEYAVALFED
 VESVGRPEVAGHSGTGLIEIAAHVAVLDEACALVADRATLMQATPGAMTALAE
 TEDIILAHDLDTVAIAVNPQSGVYSGEBAEELTATFARGKTRHREAVSFPSD
 PRMGMDARIYVAEGITRYAPRPLVSDNGRRADDEVCAGWYHREAVRPSD
 CVRTLROKATTELEGSDDLTMABEDTGLDHDHAEYPLMRAGARELAATLAR
 LQVYQVDMABTLAGGARTRIDPYTAQVHYWQPLTEPPAALAAVADDOQJMAAR
 VERDEAEILDLIGLESDOTPLDLSLTATFMSRNGENOKHLDTLRVRYMPLSRP
 TAPYDSTLMLVASDIAADAPODLGLDGLASGARARLLDGLSDADBAVLAERL
 ARTADVAATQVLYSTLDEBRADDCPLRLGLALVLYAONATGAGMLTAEERL
 AVSTNPADPTTHYQOAMNGELGVALHEPRMGLVDLPQYDTRAQORLAGLANK
 VAVDQEOUVALRATQVSGRRLVHTVEYLAETAEATAGVYLITGTCGLAEVAERL
 RARAOHLYTTSRGPDAEPALRLELQEPPEVSAVCADVADLAVALTALPEEL
 PLTGVNTHAGVCHYGLDPLSTAEFAGCTLAKLAGAHMLADLADRLPEFLFSGS
 LAGVSSNOSAGYAAANYDALAHFRAGLAASTVANGPAMALMADVDSLEERL
 OGCLIDPAPAMELBRAYRQVYTYVADYQVXAPLFSARSALITAGLPEYRAL
 ADERTKODATKSEVYTRARALAEQTLTLDLITRESAVLYSHSADVAPGRAR
 RDVDESLTAVELKRRKAGATYGLSTEWYRPLPPLAEYLRLEIGVLYEVAQV
 ATGADDEPLTALIGMCRPFQVSGSEPLADYAGDGLASEFVPMRQMGHLEDD
 PDRCQTYSTOGLLEHDEDEPTFGFISPRALVADPOORLLETTHESERAGITRD
 ELTSLATLHFNKSSVOEYGLAGDCEBHTMGSSPYLSGRSLSYNGLEGPAYVD
 TGACSSIVLAHACQSLRNGESMLAVAGCATITMTPNPPLIRVATLAKORCKATS
 DDAGMILAGVCEVYLERISDQNRNGHYVALUKGSLANDGASNGTLAGQSOOR
 YIROTALNARLAGEIDLAHEHGTGLPDLDEALQATYGRDPELSALTSVYS
 VITGTOAGACIASYIKVYALRISEPLTLHADASHVMSGCVRLTRDOLARMPET
 LRPRRAVSSFGISGNHLYLEQADYDPLTAEERPAVYTAAGVPMWYTRXSAA
 ALRQOERLAAHYETGALPPAAGPLDQTLSTVSRAPREHRAVVPAPGAPLALAL
 AVARDGSPVAVAGVADBERYTFEPQGGOSQWQMSQULDEKSAVPERIARTEALAL
 AELPMDSLVDYLGKGVASLEVDYVQVQASVAVSYLALRVEGGOVPLDVAHSGV
 ELIAAVALSLRDGARVAVLEASQILGALRGKMSVALSYDEPLREPEEGSV
 SVAAVNPRESVYAGPEALDILHAETLADYDRIARIVASVSHOVEDHEELLEV
 LAELAPTSEPPFTYTGDMILDTAMMDGYFRULRGVRLPDAVADOLLAETRAEV
 EVSSHPLVSAVVEALIDEAGVPAVAGTLRROQGTDRLLTSAAEVPERAVYDWMAL
 FEGGASARIDPLTARCHEHDSIVPAPVAVAPADKAETVATEDDSALTLAAGL
 TDEDSVAVIDPLTMSRRARDETSIVDARYVAAKMPGLGTLPRLSGHMLTAGOG
 DHDVAGALFETYGAEVRLYVDEECVDAVAREFLAGAEVNTGVSYVALAEPTDAV
 PGTSLYGLTATLALLOLDAETIDPAWALTIRGAVSTGRADELTAPVQAOYTCITMT
 LAESTHPRMGWGTIDPLPADARAOQAAVAVSGLASDSQULAHPSEYTRIVRAAR

Query Match	49.3%	Score 49.8	DB 1	Length 123	80
Best Local Similarity	68.3%	Pred. No. 0.043			
Matches 69	Conservative 0	Mismatches 32	Indels 0	Gaps 0	
QY 1	CGGTGCAAAATCGCCCGATGATAGGGGGCCGAGGTGTTGCCACGGCGAGCCCGTCCAACT 60				
Db 39095	CGATCCAAATCGCCCGGATGTGGTGGCGAGGTGTTGCCACGGCCAGTGAAGGCAAGT 39154				
QY 61	GGCGACGGGTTCCAGGCCATGGGCGTGGCCGCGCACCATC 101				
Db 39195	GGGACGTGCTGCTCGCTCCCTGGGGCGGCGCCGACGACACATC 39195				
RESULT 12					
AX211739					
LOCUS	AX211739	125401 bp	DNA	linear	PAT 06-SEP-2001
DEFINITION	Sequence 35 from Patent WO0159126.				
ACCESSION	AX211739				
VERSION	AX211739.1	GI:15523950			
KEYWORDS					
SOURCE	Streptomyces noursei.				
ORGANISM	Streptomyces noursei				

[illegible]

```
/strain-"Sg a15"
/db_xref="taxon:41"
complement(14978..6231)
/feature="ORF8"
/codon_start=1
/transl_table=1
/product="cytochrome P450 monooxygenase"
/protein_id="CAD13077.1"
/db_xref="GI:19572310"
/translation="MLRVSASRPSPANPAAPDLARDADAFVADPPFLYRAMBERO
PVHHPDGLFLTRRADJCIYRDTRTSSDKRAEFGAKYGADAPLEHHHTTSLVNDP
GLHDPTEERGLRGWILALIGALEPRLTAEOEGCNEAVTEFLDYRLIYAOBRAR
GGDPAIDVLRILGIESIKERTELTHOCVILLNAGHETTINLIGNALIELARPD
BRARLLRPAIIPTAIVEIYLYESSNOLGNRVANDREIGGVAAPPTGELTICGAN
RDPAREDEPHDQROINRIILAFGAGNHCAGMNLARIENRIALALARPPOVALT
APYRARARRRGFTAMITRLGSLR"
6369..7730
/feature="ORF7"
/codon_start=1
/transl_table=1
/product="protein kinase"
/protein_id="CAD1078.1"
/db_xref="GI:1957311"
/translation="MEELISEALPDGSGVGFWRIEGRAGYGTGAVYRARHVRICPR
SVALKARYPNDRERIELLERIRHANVPRLLGHWGRAGPQGEAPRYVMOVMEG
VRLQWNERLRTAREVILRVLAQVARSALAEATHMRGILRDVKGENVILSPGHSILMD
FGCGTWEGSAAPLJTGILAPGTAKARSPALRHHNRNEADAHYQATPADVYALGVT
AYRLCTGIYPPPATPEPSCGVGEDERTQEVKKPPSQVRPLVPALDELILMLRDNPKER
GRAEELAEEMALERAAGEADEEPNPNHSSVPSBITGRPVSVOREMGWPGVALIVA
VGLLWGSGSLNTGRRVAPASMTGCGGVADAEVALPVSSARAREPSPGLNIDML
KEPLGGRRPPCPRQIISIGGCWLEVPQSSPGCESYEWKACYSVPAAPPRPTS
DPO"
complement(7849..3577)
/feature="ORF6"
/codon_start=1
/transl_table=1
/product="methyl transferase"
/protein_id="CAD19079.1"
/db_xref="GI:19572313"
/translation="MSQNIYDDPSFFESYSOLARSLKGLAGAPEMPALRALLPYKDL
RVLDGCGEFGWCHARKHDAHVGLDLSKMLTRAKETSGAITYQOADELVLV
PEGAFDLAFSSLAHYVELLEPLATVIRGLTPGGMVFVSTIEHPITMASRQPEWHTGP
EGHRTWPVDSYOLGSKRTIHWLAEEVYVYHRTLTGTLVALLBOGTTLHIEWGPTDA
OLSAPELAEERDPMFLVSAOR"
complement(8622..4879)
/feature="ORF5"
/codon_start=1
/transl_table=1
/product="acyl-CoA binding protein"
/protein_id="CAD19080.1"
/db_xref="GI:19572313"
/translation="MALEIDERSAQRVAKLKTSPSNDRLLELISLKQATEGVQOK
RPGILDKERAKYAWGHIKGLGRESAQOYVALVERLLRG"
8961..10070
/feature="ORF4"
/codon_start=1
/transl_table=1
/product="potassium channel beta chain"
/protein_id="CAD19081.1"
/db_xref="GI:19572314"
/translation="MYVLLJRSQSLRWILNGSAHRVAFGCHARRPMHFRHLRGS
GLWSEISYGMNLTHGSOVREASAVACVRAALDEITTFPDADVAGRAEVEVIGRAL
OGORREGLEITTKVYPTGKPNDRGLSRKHILIESIHGSLRKLREYDLYOAHRYDT
ETPLEETMOAFSDVROGKHYIGVSEMTADLRGCVLSROMQTEVLSQPOYSMLR
RVIEAEVYASOELEGVGOVYSPITAGVLTGKYRGOKPPGSRATIDDEKGSATFKRP
LNDVLTVPVQOLTVADEAYILSLAOLAVAVYIOWNPVSSALIGASRPQVAKENYKAAK
VKLSSEAMRHITDLGPIVIRNPKKTLNSRKRP"
10438..10862
/feature="ORF3"
/codon_start=1
/transl_table=1
/product="ribosome binding factor"
```

```
/protein_id="CAD19082.1"
/db_xref="GI:19572315"
/translation="MTSEKORRRRAPRRGTSSLSFEASOILPAHHLRVOSTLROEVA
RLFRGELSDPLLEGVSIVSEFLTPDGRNARIGFLVPEYEGGSAPLEALVRSQFI
RAQLALGLNLRKVRPHLRFCVGAAPGFSIPEDGSABEGBS"
10859..12016
/feature="ORF2"
/codon_start=1
/transl_table=1
/protein_id="CAD19083.1"
/db_xref="GI:19572316"
/translation="MRPRRIGAGVAPILAMARAVPAGAVAKOLLLAROPYLAENYA
AMPVHVAEGVAVGVTEFATERHIVYDGLGGCVASAREDPFAMSLGRVELRLLA
AFARAVPYGDVHRGRLSLPRGIDIGLSTHRLQRMERLVPKHLCTGGCNGLEEL
DRDEGALMILLHNGSRGVCAITAHHLRVAVSGESLPLGLRTDSGGAACIADPEL
ACQBARANRPTLIORAVEVLAGIQQGVYDLSRVYHHNNHYAASHRGRTLYARKKA
VGIYAGCGGLIPGSMGTSYLVAKGKPOACSGKAGVMTREARERTRALAH
ALRRVYDEGSVMTLVERAPRAYDITVELEDEALVPLRLPLIAYLNG"
complement(12082..13974)
/feature="ORF1"
/codon_start=1
/transl_table=1
/product="cellulase"
/protein_id="CAD19084.1"
/db_xref="GI:19572317"
/translation="MKNSLLLMLAVLPLXGTALARPMANAEPHGEPEENAGLTA
AOCNRPVPLSTRGYIYDRCEGERLKSVMNFGASDQLEVVGGIDKOKLSDVLAQMA
MGFNSVRLPFSNNMLHPDNCKPAQORCLEKYSVTCIDSTRKPELINTKTALEYDVAVA
ELTRQGLVYILNNHTTSMCCGMDNGFQDSQALQRMDDVMMLAARQGNKMWAG
ADLNRENVDPDLDSPNMGMRQHMWMAAQTMNLILRTPDILLIVEAVNMGMLDGG
SRPDLKPYRQAPVALLRQDKLVYVHYNGYTPQNSQSLSGSGPYSDMDKPTLYGLT
DQEWGFLAANAQAVTAPYWMSEFGVYNEQANARAFSNLADYLIEKDWMAWYALN
AAKLONSVOGEDETGYLMSYPDMNGVSGPQSGTGLGRILGADGRTGADVAPFLP
MTVLKDGSSYTYVDNNSLDSDSHSGAKISCPGTGYMVGVSANFTLLCNAGVAPAO
OGTSGYOTITQIEVPLRYPGDMANOSIKFEDPTQFAFVAGISTANPFLERAGLCEON
TGCIPLNKSRAKNMGWDQIRASSGSGGMLVETKQCCADNOXITGLIHRRSGLADYS
VLCSN"
14345..21466
/feature="stia"
14345..21466
/feature="stia"
/function="PKS"
/codon_start=1
/transl_table=1
/product="stia protein"
/protein_id="CAD19085.1"
/db_xref="GI:19572318"
/translation="MTQLASRLGVSGSESIDPRERSYQICDSIRTKSLAIFGLIKR
PLSTPLANTITVTEALSHHLAGNDVSGGERADLYQDQDEPILITMAGRFKAPVY
ESFNSLNDGVDALITEVPPERMDIARTSVGLKSPGKANTMGFLLQVQDFLPGF
ISPREALOMDPOORLALBELSWEALESAGVAVDSLMDSRTEGVFKGAMSDTAKTLRGL
GLIOPHTYTGQDSSITARVSYFLGLRQPSLTVPKASSALVAHMLCQSLRNEAM
ALAGGVNLLSPSSVAMSOIGALSPGRCAAFVAPRANGVYRGGGVVVLTPLSRAT
ANGYPLVCIRGSAVNDGFSNGLTAPRPAKQEVILHDELAASGIAGLMVYSTIQR
GTVLGDIPEANALSAVLGVNRPAPRPLTGSVKNINIHLEASGIAGLMVYSTIQR
AIIPNLHEQPNPYIRFEEELRLVAPKSLHPNPNNGSEVLAIGISFEGGNCNCHLVLEA
PSPGSANITAPLAETPEBELSLARQADNAVSVSOVELADICLSVARAREGSORLA
LIARSPALIAQVDFLAGERKPGILRKEVYLSGSSQKPEVFVSGTGSQWPMQMGRELIR
TEPVRGALIERCDALRLSGMSMLELSAEGASRKECSGLQTAIFAIOVYALALM
RSMQIERPAGVYGHSIGEVAANHAGIINLEDAAPYIHLISLMKRLSGQAMIVYDS
AEEANRLIREMERHVVVAAACNAPASTVLSGARVEIIEIGALEIRGTAIAHVRIDTS
HSPAMDPEMGELAVLAGLSPRAVAVSVTVTAFAFKAECTAAVYVRNIREVLS
QALTKLEQDRLRLLESPIHPILARSVOCILKSGRGQVLAISKRRENEGACAREEL
GRLTLEGVSVMQWLYPEAKRVFLPGLGTGSGIDALRGSLARPLPSATHTALEALRE
ARALQASIAKHPVDRLDLCYSAVGVSVYHNRPLATDGSRAVLEHLDPLKLESPP
GRIGVCACTGEPKVVVPEPGOSOMQGOALRLQRPVREALIEEDDAIRSGAGLS
LIDALQDEPHSRGSIETVIOPTFAILEVLAALMRGVYVPAVAVESHGNETAAACV
AGALRTDARVTCERSOLPRVAGCGGMLVETSLBOAREMISRKRGVRSIGASNP
RSYVSGDAIATLEILRIELNOCGRQRYVAVDSPOVDPLKDLRLVYISPTAPA
ATVYLSYGEPTIEGARMDASVMRILREPVIRGQALIERLAGCHGVYVETISPHPL
LPSTIDOTLRKRSKGYLPLSRREPERAVILBEGTCLTFGGGQYVDRGLHPLGGHYV
RLPHUPMORQRFMLDMDDVAAAGAASVGAAMPGGGSHPLGLGHPLESLAHPACHTVE
NKLRSASVSTLDHRVGGVYVLPOTSTIEMALAAANGSGKTVSLRDVLEKMSFL
```


8
.
.
.
/

—

—

/


```
; Sequence 8, Application US/09789836
; Patent No. US20020082204A1
; GENERAL INFORMATION:
; APPLICANT: BRIGHAM, KENNETH L.
; APPLICANT: STECENKO, ARLENE A.
; APPLICANT: SEALY, LINDA
; TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
; FILE REFERENCE: N-6977
; CURRENT APPLICATION NUMBER: US/09/789,836
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/183,584
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1500
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-789-836-8
```

```
Query Match      29.3%; Score 29.6; DB 10; Length 1500;
Best Local Similarity 61.8%; Pred No. 4.9;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 16 TGCAGCACCACCGCGGTGTCAGCCGACAGCGCGCGCGCGGTGCGG 75
DB 739 TTGGCGGGCGCGCGCGCGCGCGAGAGCAGCGCGCGCGCGCGGTGCGG 680
OY 76 CGCCTCGAGGTGCTGG 91
DB 679 GGGCTCGCGGTGCGCGG 664
```

```
RESULT 3
US-09-923-876-2938/c
; Sequence 2938, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laiguel, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Shetman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 2938
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US:0020013958A1 700161420H1
US-09-923-876-2938
```

```
Query Match      28.5%; Score 28.8; DB 10; Length 239;
Best Local Similarity 60.0%; Pred No. 9.1;
Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 22 ACCCAGCCCGCGGTGTCAGCCGACAGCGCGCGCGCGCGGTGCGGCGGCG 81
DB 219 AGCCGGGTCTCCATGATGTACAGCGGACGAGCGGAGCCGACCGACCGCGCG 160
OY 82 GAGTGCTGAGAGCGCGCG 101
DB 159 GAGTACTGACAGCGCGCGCG 140
```

```
RESULT 4
US-09-925-301-122/c
; Sequence 122, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 122
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (41)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (218)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-122
```

```
Query Match      28.3%; Score 28.6; DB 10; Length 656;
Best Local Similarity 58.3%; Pred No. 9.5;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 16 TGCAGCACCACCGCGGTGTCAGCCGACAGCGCGCGCGCGGTGCGGCGG 75
DB 271 TGCAGCACCCTCCAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 212
OY 76 CGCCTCGAGGTGCTGAGAGCGCG 99
DB 211 CCCCTGAGGTGCGCGCGCGCGCG 188
```

```
RESULT 5
US-09-864-761-12628/c
; Sequence 12628, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
```


;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 12628
;; LENGTH: 531
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL121673.27
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.68
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.75
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
US-09-864-761-12628

Query Match 28.1%; Score 28.4; DB 10; Length 531;
Best Local Similarity 58.1%; Pred. No. 11;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 AGACGAGCGCGATGAGACGACCGCCCGCGGTGCGTGCAGCGAGAGCGCGCC 61
DB 388 AGCCACTGCGCCCTGCGAGGCGCCAGTGGGCGCGCCGCCACGACGAGATGAGGAG 329
QY 62 CCGGCGCGTGGCGCGCTGAGGTG 87
DB 328 ATGGCAGTACCGATCCCGGAGGTG 303

RESULT 6
US-09-933-814-1
;; Sequence 1, Application US/09933814
;; Patent No. US20020058798A1
;; GENERAL INFORMATION:
;; APPLICANT: WALLACH, David
;; APPLICANT: BOLDIN, Mark
;; APPLICANT: VARFOLOMEYEV, Eugene
;; APPLICANT: METT, Igor
;; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/APOL RECEPTORS
;; FILE REFERENCE: WALLACH-16B
;; CURRENT APPLICATION NUMBER: US/09/933,814
;; PRIOR FILING DATE: 2001-08-22
;; PRIOR APPLICATION NUMBER: 08/860,082
;; PRIOR FILING DATE: 1997-08-19
;; PRIOR APPLICATION NUMBER: PCT/US95/16542
;; PRIOR FILING DATE: 1995-12-14
;; PRIOR APPLICATION NUMBER: IL 112022
;; PRIOR FILING DATE: 1994-12-15
;; PRIOR APPLICATION NUMBER: IL 112692
;; PRIOR FILING DATE: 1995-02-19
;; NUMBER OF SEQ ID NOS: 2
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 1
;; LENGTH: 1701
;; TYPE: DNA

;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(768)
US-09-933-814-1

Query Match 28.1%; Score 28.4; DB 10; Length 1701;
Best Local Similarity 60.3%; Pred. No. 9.8;
Matches 47; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 24 CCACGCCCGCGGTGCGGTGACGACGAGCGCGCCCGCGCGTTCGCCGCTCGA 83
DB 66 CCGGCGAGGCGCGAGCGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125
QY 84 GGTGCTGGAGAGCGCGCC 101
DB 126 GCGCGTTCGACACCCCGC 143

RESULT 7
US-09-824-134-1
;; Sequence 1, Application US/09824134
;; Patent No. US20020082401A1
;; GENERAL INFORMATION:
;; APPLICANT: WALLACH, David
;; APPLICANT: BOLDIN, Mark
;; APPLICANT: VARFOLOMEYEV, Eugene
;; METT, Igor
;; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/APOL RECEPTORS
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BROWDY AND NEWMARK, P.L.L.C.
;; STREET: 419 Seventh Street N.W., Ste. 300
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: United States of America
;; ZIP: 20004
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/824,134
;; FILING DATE: 03-Apr-2001
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/860,082
;; FILING DATE: <Unknown>
;; APPLICATION NUMBER: IL 112022
;; FILING DATE: 15-DEC-1994
;; APPLICATION NUMBER: IL 112692
;; FILING DATE: 19-FEB-1995
;; APPLICATION NUMBER: IL 114615
;; FILING DATE: 16-JUL-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BROWDY, Roger L.
;; REGISTRATION NUMBER: 25,618
;; REFERENCE/DOCKET NUMBER: WALLACH-16
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 628-5197
;; TELEFAX: (202) 737-3528
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1701 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..768
;; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-815-242-7925

Query Match
Best Local Similarity 58.3%; Pred. No. 13;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 18 GACGACCCACGCGCGGTGGTGACAGCCGAGCGCGCGCGCGGTGGCGG 77
Db 1215 GATCATCGAGCGCCGAGGCGGTGACCGAGCGGTGCGCCCATCGAGAGATGCT 1274

QY 78 CCTGAGGTGCTGAGAGACCGCGC 101
Db 1275 CGACGAGTCTGCGCGCCACGCG 1298

RESULT 11

US-09-789-836-10/c
Sequence 10, Application US/09789836
Patent No. US20020082204A1

GENERAL INFORMATION:
APPLICANT: BRIGHAM, KENNETH L.
APPLICANT: SNECENKO, ARLENE A.

APPLICANT: SEALY, LINDA
TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20

FILE REFERENCE: N-6977
CURRENT APPLICATION NUMBER: US/09/789,836

PRIOR FILING DATE: 2001-02-20
CURRENT APPLICATION NUMBER: 60/183,584

PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 33

SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 10

LENGTH: 1739

TYPE: DNA
ORGANISM: Rattus sp.

US-09-789-836-10

Query Match
Best Local Similarity 27.7%; Score 28; DB 10; Length 1739;

Matches 46; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 16 TGAACGACCCAGCGCGGTGGTGACAGCCGAGCGCGCGCGGTGGCGG 75
Db 761 TTGGCGGGCGCGCGCGCGCGCGCGAGAGCGCGCGCGCGCTTGGCTGGCGG 702

QY 76 CGGCTCGAGGTGCTGG 91
Db 701 GGGCTCGCGCTGCCG 686

RESULT 12

US-09-925-299-166/c
Sequence 166, Application US/09925299
Patent No. US20020055627A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102

CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1556
SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 166
LENGTH: 2061

TYPE: DNA

ORGANISM: Homo sapiens
US-09-925-299-166

Query Match
27.7%; Score 28; DB 10; Length 2061;

Best Local Similarity 59.7%; Pred. No. 12;
Matches 43; Conservative 2; Mismatches 27; Indels 0; Gaps 0;

QY 27 CGCCCGCGGTGGGTGACGACGAGCGCGCGCGCGCGGTGGCGGCTGAGGT 86
Db 74 CGCGCTGCTGAGGTGACGAGGTGACGACGCGCGCGCGCGCGCGCTTCGCGCT 15

QY 87 GCTGAGGAGCGG 98
Db 14 GCTGCGGTGCGG 3

RESULT 13

US-09-923-876-1480/c
Sequence 1480, Application US/09923876
Patent No. US20020013958A1

GENERAL INFORMATION:
APPLICANT: Laligudi, Raghunath V.
APPLICANT: Kamigaki, Laura Y. (Ito)

APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING

FILE REFERENCE: PL-0012-1-CON
CURRENT APPLICATION NUMBER: US/09/923,876

PRIOR FILING DATE: 2001-08-06
CURRENT APPLICATION NUMBER: 09/298,329

PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085,331

PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 6332

SOFTWARE: PERL Program
SEQ ID NO 1480

LENGTH: 224

TYPE: DNA
ORGANISM: Zea mays

FEATURE:
NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. US20020013958A1 700158889H1
NAME/KEY: unsure

LOCATION: 96, 116, 148, 173, 189, 211
OTHER INFORMATION: a, t, c, g, or other

US-09-923-876-1480

Query Match
Best Local Similarity 27.5%; Score 27.8; DB 10; Length 224;

Matches 47; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 22 ACCGACGCGCGCGGTGGGTGACGACGAGCGCGCGCGCGGTGGCGGCTG 81
Db 107 AGCCGCGCTGCTGACGAGGTGACGAGCGGAGCGGAGCGCGCGCGCGCGCGCG 48

QY 82 GAGGTGCTGGAGAGCGCGG 101
Db 47 GAGGTACTGACGCGCCACGCG 28

RESULT 14

US-09-764-887-416
Sequence 416, Application US/09764887
Patent No. US20020042096A1

GENERAL INFORMATION:
APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA113

CURRENT APPLICATION NUMBER: US/09/764,887
CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 658

SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 416

LENGTH: 464

TYPE: DNA
ORGANISM: Homo sapiens

US-09-764-887-416

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds

(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_28200_28300

Perfect score: 101

Sequence: 1 gagacgcgaacgcgatgac.....gaggtgtgagagcgcgcgc 101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_hlg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hlg_hum:*

31: em_hlg_inv:*

32: em_hlg_other:*

33: em_hlg_mus:*

34: em_hlg_pln:*

35: em_hlg_rtd:*

36: em_hlg_mam:*

37: em_hlg_vrt:*

38: em_sy:*

39: em_hlgo_hum:*

40: em_hlgo_mus:*

41: em_hlgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1 AF217189	AF217189 Sorangium
2	101	100.0	68750	1 AF210843	AF210843 Sorangium
3	101	100.0	68750	6 AR193029	AR193029 Sequence
4	101	100.0	68750	6 AR199551	AR199551 Sequence
5	101	100.0	68750	6 AR199559	AR199559 Sequence
6	101	100.0	68750	6 AR199567	AR199567 Sequence
7	101	100.0	68750	6 AR201097	AR201097 Sequence
8	101	100.0	68750	6 AR208671	AR208671 Sequence
9	101	100.0	71989	6 AR172664	AR172664 Sequence
10	36.8	36.4	80966	2 AC119719	AC119719 Rattus no
11	36.6	36.2	155406	2 AC105514	AC105514 Rattus no
12	35.6	35.2	92600	2 AC128756	AC128756 Rattus no
13	34.8	34.5	134366	2 AC119661	AC119661 Rattus no
14	34.6	34.3	152355	2 AC109839	AC109839 Rattus no
15	34.4	34.1	154607	2 AC008193	AC008193 Drosophila
16	34.2	33.9	896	11 PM12F6G	PM12F6G Penicillium
17	34	33.7	190459	2 AP005419	AP005419 Oryza sat
18	33.8	33.5	83383	2 AC095331	AC095331 Rattus no
19	33.8	33.5	176695	2 AC126065	AC126065 Rattus no
20	33.6	33.3	2936	1 AF220005	AF220005 Rhodospirillum rubrum
21	33.6	33.3	111084	9 AC006486	AC006486 Homo sapiens
22	33.4	33.1	719	9 HSA324403	HSA324403 Homo sapiens
23	33.4	33.1	1266	1 MFA251337	MFA251337 Mycobacterium tuberculosis
24	33.4	33.1	69160	2 AC094374	AC094374 Rattus no
25	33.4	33.1	84464	9 HS782D21	HS782D21 Human DNA
26	33.2	32.9	97967	2 AC103236	AC103236 Rattus no
27	33.2	32.9	181400	2 AC111558	AC111558 Rattus no
28	33.2	32.9	185575	2 AC125578	AC125578 Rattus no
29	33.2	32.9	262549	2 AC113623	AC113623 Rattus no
30	32.8	32.5	99539	2 AC097740	AC097740 Rattus no
31	32.8	32.5	172435	8 OSJN00163	OSJN00163 Oryza sativa
32	32.6	32.3	106946	2 AC118298	AC118298 Rattus no
33	32.6	32.3	125481	2 AC120974	AC120974 Rattus no
34	32.6	32.3	143330	2 AC094396	AC094396 Rattus no
35	32.6	32.3	154587	2 AC127744	AC127744 Rattus no
36	32.6	32.3	190457	2 AC119326	AC119326 Rattus no
37	32.4	32.1	28731	2 AC121461	AC121461 Rattus no
38	32.4	32.1	76292	2 AC128848	AC128848 Rattus no
39	32.4	32.1	139629	2 AP003809	AP003809 Oryza sativa
40	32.4	32.1	157448	8 AC131139	AC131139 Rattus no
41	32.2	31.9	1156	8 AF342809	AF342809 Zea mays
42	32.2	31.9	14675	1 AE007017	AE007017 Mycobacterium tuberculosis
43	32.2	31.9	37750	1 SC6D7	SC6D7 Streptomyces
44	32.2	31.9	39150	1 MTCY21B4	MTCY21B4 Mycobacterium tuberculosis
45	32.2	31.9	46202	2 AC096789	AC096789 Rattus no

ALIGNMENTS

RESULT 1
AF217189
LOCUS
DEFINITION
AF217189 58733 bp. DNA linear. BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189.1 GI:7453554
KEYWORDS
ORGANISM
Polysangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Mycobacteriales;
Mycobacteriales; Sorangiales; Polysangiaceae; Polysangium.
REFERENCE
1 (bases 1 to 58733)

Pred. No. is the number of results predicted by chance to have a

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115995
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Zierman, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20230308
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submission
Submitted (16-DEC-1999) Korean Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
JOURNAL Location/Qualifiers
FEATURES
source 1. 58733
/organism="Polyanthus cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
<1. 992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453355"
/translation="VRGIVEAFADFGVPRVLLYDNKNAVYERHODAIRPHPTLAL
SDYREPRPVAVARGNKGREARIRYVRGEFFERARADIGDNRQATMTSSAL
DSWYREPRPVAVARGNKGREARIRYVRGEFFERARADIGDNRQATMTSSAL
DPTRETVYADLSQVRIADGNOIVATYHRSMDROQIQIPHELDYDEKRAHRR
GIDRLARARSSQARIRIYAERGDVNGSIALNLDLVGADELAIVERLDTI
HIGAVQVYDNRKSERHLPPIPIPTKEHALVYTPHSTITDLKADPTP
969. 1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62379.1"
/db_xref="GI:7453356"
/translation="MTDTPTETKDKLSLGLGLACMEQLADKRWLEKEVLAIEERE
RKRSLERLKNRYAAAF(PMTDFDSSWPKIDREAVVDLYDSRADLLFEVYTRKYD
AOKPLLTSTNKAFAWMGQ/PPHACVYTLVDRLVHRAVEIEIEASYRLKEKELNAT
RTKQRTKKH"
1998. 6263
/gene="epoA"
1998. 6263
/gene="epoC"
/note="type 1 polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62881.1"
/db_xref="GI:7453357"
/translation="MEEDSSAIVIGSGRRPGGARDIDFEFRNLRDGTFAVORSE
GVAERGDNRARLDHPPESSGREKREVEAAGDRPFLRLEIDEGRLVLRVYERAPG
IGVEFLAVDAAGLSTFNQVIALGMPDDLPKPNPILIDEGCCACAGIYAVGVGAGV
VGOPIYATASAGAPATHVTTSALVYPRPQASATPEAAMPYVLTAWAYALDIAIAP
GPRVLIHAATGCVGLAANOMAOHVGAATATGPEKRAYLIESIGVYDSRSDRPV
ADVPATGGEVDVYLNSTSCETIKRSNLTPSGRPEVEIGCRDCAVNOAGTPEIR
NISEYLDRCGMIRPARVYALFPEILGCTIAAGVPPPIATITLIPAYADPREMO
AOLIKIVITLIDCEVOIRIPTHAAGSGDRDLIDLASAARAPAAAEAFRTMO
VQOYVPEITKGAENLFTRLDMOSLMAVELNRIEASIKLNLSTFTSTSPNLI
QNLDLNALSLEKVAENLRAGVSDPEVSSGADQDMETAL"
6260. 10492
/gene="epoB"
6260. 10492
/gene="epoC"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453358"
/translation="MTINOLNEHQVKLADGERLQIQAPKNALPNILARISEH
KSTILMLRORPAESIYVAPAEHRAVPEPLDIOGSYMLGRTGATTVPSGIHAYRED
CTDLVVARLSRAFRVAVAHDMRLAHTLPDMQVIEPVYADIEITIDRGDRSTREA
RLVSLDAMASHRIYDTERPELVAVAVRLEDEQOTLVSLIDLVNDLSLSTIFKDM
SFYEDPETSILPVLESLRDYVYALSRKSEAHORSMDYMKRRAVELPPPLMKAD
PSTLREIRPHREOMLPSDSWGRLOKORGERGLPTGYILAAFSVIGRMSASPELT
NITLENRLPVHPRVNDITGDETSMVLIDITDRKSPQORAKRIOEOLMEAMDHDVS
GIEVOREARVIGIOGALFPVLTSAINQOYVYTSIORIGTIPVYSTOTPOLLDH
OLXENDGDIYLAAMDVDFPDDILDAEVAATLRLTEPESMORCSIPAPLE
ASASNNISLSEHTLHGLFPAARBDLPMLAVASAKTITTEESRSGISAEHL
BOGARPNITVAYAMKRGEOYAVAVAVESGAAYVPIADIPAEIHYLIDHGEKIV
LTOPMLDLSMPGICORLVDAVDEBDGLPMPQITDASLAVYVYSGSTLKP
CYMIDHRAVYNIIDINREPEIGCDPRALATIESFQSYVDYDRIITAAAGIYVPA
SKLRPRAHMANIIEBEKYVWNSYALMMLVHEHSGEDSLARSLTSLSGMIVP
GLPEQAIKROGVSYISLQGTAKSINSGIYKRVAVDSMSIPIGRRLNQTPIYLD
EALREPMVWNGQILIGGYLALGTMRDEKRNPLVHPEVEGLITGDGRLIPD
GNIERPMEDNOIKILGTREVEGLIEETLRKNSPLVHPEVEGLITGDGRLIPD
EETRRAAEODNSLARKRYRFLLEAPRIVERGRSLCSSEVPDGAALPKRPVSAG
TGLVPRAGLDVYARKRSYRFLLEAPRIVERGRSLCSSEVPDGAALPKRPVSAG
STYVQTIYAVAKSGRIIEGVDEGFYIYHPEHMLLVSHGLENGVAVQNPDEPEEA
FCLFVGRIDALIESLYGSLREFCLEGGYMLMEABOSCNIGVAVGQDEFOVR
FVLDIRHSDVYVHGMGLGVDRQFOVCTLGDSSPRKATISGAPGHDADILRD
FLTRKLPKEMVYVEVDEDALPDTSGNGKRAKALERRDTSSPRSGHTAPADALEI
LVAAYRVSGLEVEVLAQSFVDLGATSIHIVMRSLQKRLDREIATIELFOYPLIGS
LVAAGLRSDKLEDRPNMODRVEARRKGRRS"
10639. 16137
/gene="epoC"
10639. 16137
/gene="epoC"
/note="type 1 polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453359"
/translation="MEEDSSAIVIGSGRRPGGARDIDFEFRNLRDGTFAVORSE
GVAERGDNRARLDHPPESSGREKREVEAAGDRPFLRLEIDEGRLVLRVYERAPG
IGVEFLAVDAAGLSTFNQVIALGMPDDLPKPNPILIDEGCCACAGIYAVGVGAGV
VGOPIYATASAGAPATHVTTSALVYPRPQASATPEAAMPYVLTAWAYALDIAIAP
GPRVLIHAATGCVGLAANOMAOHVGAATATGPEKRAYLIESIGVYDSRSDRPV
ADVPATGGEVDVYLNSTSCETIKRSNLTPSGRPEVEIGCRDCAVNOAGTPEIR
NISEYLDRCGMIRPARVYALFPEILGCTIAAGVPPPIATITLIPAYADPREMO
AOLIKIVITLIDCEVOIRIPTHAAGSGDRDLIDLASAARAPAAAEAFRTMO
VQOYVPEITKGAENLFTRLDMOSLMAVELNRIEASIKLNLSTFTSTSPNLI
QNLDLNALSLEKVAENLRAGVSDPEVSSGADQDMETAL"
6260. 10492
/gene="epoB"
6260. 10492
/gene="epoC"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453358"
/translation="MTINOLNEHQVKLADGERLQIQAPKNALPNILARISEH
KSTILMLRORPAESIYVAPAEHRAVPEPLDIOGSYMLGRTGATTVPSGIHAYRED
CTDLVVARLSRAFRVAVAHDMRLAHTLPDMQVIEPVYADIEITIDRGDRSTREA
RLVSLDAMASHRIYDTERPELVAVAVRLEDEQOTLVSLIDLVNDLSLSTIFKDM
SFYEDPETSILPVLESLRDYVYALSRKSEAHORSMDYMKRRAVELPPPLMKAD
PSTLREIRPHREOMLPSDSWGRLOKORGERGLPTGYILAAFSVIGRMSASPELT
NITLENRLPVHPRVNDITGDETSMVLIDITDRKSPQORAKRIOEOLMEAMDHDVS
GIEVOREARVIGIOGALFPVLTSAINQOYVYTSIORIGTIPVYSTOTPOLLDH
OLXENDGDIYLAAMDVDFPDDILDAEVAATLRLTEPESMORCSIPAPLE
ASASNNISLSEHTLHGLFPAARBDLPMLAVASAKTITTEESRSGISAEHL
BOGARPNITVAYAMKRGEOYAVAVAVESGAAYVPIADIPAEIHYLIDHGEKIV
LTOPMLDLSMPGICORLVDAVDEBDGLPMPQITDASLAVYVYSGSTLKP
CYMIDHRAVYNIIDINREPEIGCDPRALATIESFQSYVDYDRIITAAAGIYVPA
SKLRPRAHMANIIEBEKYVWNSYALMMLVHEHSGEDSLARSLTSLSGMIVP
GLPEQAIKROGVSYISLQGTAKSINSGIYKRVAVDSMSIPIGRRLNQTPIYLD
EALREPMVWNGQILIGGYLALGTMRDEKRNPLVHPEVEGLITGDGRLIPD
GNIERPMEDNOIKILGTREVEGLIEETLRKNSPLVHPEVEGLITGDGRLIPD
EETRRAAEODNSLARKRYRFLLEAPRIVERGRSLCSSEVPDGAALPKRPVSAG
TGLVPRAGLDVYARKRSYRFLLEAPRIVERGRSLCSSEVPDGAALPKRPVSAG
STYVQTIYAVAKSGRIIEGVDEGFYIYHPEHMLLVSHGLENGVAVQNPDEPEEA
FCLFVGRIDALIESLYGSLREFCLEGGYMLMEABOSCNIGVAVGQDEFOVR
FVLDIRHSDVYVHGMGLGVDRQFOVCTLGDSSPRKATISGAPGHDADILRD
FLTRKLPKEMVYVEVDEDALPDTSGNGKRAKALERRDTSSPRSGHTAPADALEI
LVAAYRVSGLEVEVLAQSFVDLGATSIHIVMRSLQKRLDREIATIELFOYPLIGS
LVAAGLRSDKLEDRPNMODRVEARRKGRRS"
10639. 16137
/gene="epoC"
10639. 16137
/gene="epoC"
/note="type 1 polyketide synthase; epothilone PKS module
2"

AVGGLVSNAGLFPSSGRRVPLPTYPMORERYWIDTKADDAAAGDRBAGAGDEVEKG
GAVRGDRRSARLDHPPESSGREKREVEAAGDRPFLRLEIDEGRLVLRVYERAPG
IGVEFLAVDAAGLSTFNQVIALGMPDDLPKPNPILIDEGCCACAGIYAVGVGAGV
VGOPIYATASAGAPATHVTTSALVYPRPQASATPEAAMPYVLTAWAYALDIAIAP
GPRVLIHAATGCVGLAANOMAOHVGAATATGPEKRAYLIESIGVYDSRSDRPV
ADVPATGGEVDVYLNSTSCETIKRSNLTPSGRPEVEIGCRDCAVNOAGTPEIR
NISEYLDRCGMIRPARVYALFPEILGCTIAAGVPPPIATITLIPAYADPREMO
AOLIKIVITLIDCEVOIRIPTHAAGSGDRDLIDLASAARAPAAAEAFRTMO
VQOYVPEITKGAENLFTRLDMOSLMAVELNRIEASIKLNLSTFTSTSPNLI
QNLDLNALSLEKVAENLRAGVSDPEVSSGADQDMETAL"
6260. 10492
/gene="epoB"
6260. 10492
/gene="epoC"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453358"
/translation="MTINOLNEHQVKLADGERLQIQAPKNALPNILARISEH
KSTILMLRORPAESIYVAPAEHRAVPEPLDIOGSYMLGRTGATTVPSGIHAYRED
CTDLVVARLSRAFRVAVAHDMRLAHTLPDMQVIEPVYADIEITIDRGDRSTREA
RLVSLDAMASHRIYDTERPELVAVAVRLEDEQOTLVSLIDLVNDLSLSTIFKDM
SFYEDPETSILPVLESLRDYVYALSRKSEAHORSMDYMKRRAVELPPPLMKAD
PSTLREIRPHREOMLPSDSWGRLOKORGERGLPTGYILAAFSVIGRMSASPELT
NITLENRLPVHPRVNDITGDETSMVLIDITDRKSPQORAKRIOEOLMEAMDHDVS
GIEVOREARVIGIOGALFPVLTSAINQOYVYTSIORIGTIPVYSTOTPOLLDH
OLXENDGDIYLAAMDVDFPDDILDAEVAATLRLTEPESMORCSIPAPLE
ASASNNISLSEHTLHGLFPAARBDLPMLAVASAKTITTEESRSGISAEHL
BOGARPNITVAYAMKRGEOYAVAVAVESGAAYVPIADIPAEIHYLIDHGEKIV
LTOPMLDLSMPGICORLVDAVDEBDGLPMPQITDASLAVYVYSGSTLKP
CYMIDHRAVYNIIDINREPEIGCDPRALATIESFQSYVDYDRIITAAAGIYVPA
SKLRPRAHMANIIEBEKYVWNSYALMMLVHEHSGEDSLARSLTSLSGMIVP
GLPEQAIKROGVSYISLQGTAKSINSGIYKRVAVDSMSIPIGRRLNQTPIYLD
EALREPMVWNGQILIGGYLALGTMRDEKRNPLVHPEVEGLITGDGRLIPD
GNIERPMEDNOIKILGTREVEGLIEETLRKNSPLVHPEVEGLITGDGRLIPD
EETRRAAEODNSLARKRYRFLLEAPRIVERGRSLCSSEVPDGAALPKRPVSAG
TGLVPRAGLDVYARKRSYRFLLEAPRIVERGRSLCSSEVPDGAALPKRPVSAG
STYVQTIYAVAKSGRIIEGVDEGFYIYHPEHMLLVSHGLENGVAVQNPDEPEEA
FCLFVGRIDALIESLYGSLREFCLEGGYMLMEABOSCNIGVAVGQDEFOVR
FVLDIRHSDVYVHGMGLGVDRQFOVCTLGDSSPRKATISGAPGHDADILRD
FLTRKLPKEMVYVEVDEDALPDTSGNGKRAKALERRDTSSPRSGHTAPADALEI
LVAAYRVSGLEVEVLAQSFVDLGATSIHIVMRSLQKRLDREIATIELFOYPLIGS
LVAAGLRSDKLEDRPNMODRVEARRKGRRS"
10639. 16137
/gene="epoC"
10639. 16137
/gene="epoC"
/note="type 1 polyketide synthase; epothilone PKS module
2"

Query Match	Best Local Similarity	Score	DB 1	Length
Matches	101	100.0%	DB 1	58733
Conservative	0	Mismatches	0	Indels
Gaps	0	Indels	0	Gaps
1	GAGACCGAAGCCGATGACAGACCCACCGCCCGCGTCCGGGTGACAGCCGACGAGCGCGC	60		
28200	GAGACCGAAGCCGATGACAGACCCACCGCCCGCGTCCGGGTGACAGCCGACGAGCGCGC	28259		
61	CCCGCGCGGTGACCGCGCTCGAGGTGCTGGAGACCGCGC	101		
28260	CCCGCGCGGTGACCGCGCTCGAGGTGCTGGAGACCGCGC	28300		
AF210843	AF210843	68750 bp	DNA	linear
LOCUS	AF210843	68750 bp	DNA	linear
DEFINITION	Sorangium cellulosum strain So ce90 epothilone biosynthesis gene			
ACCESSION	AF210843			
VERSION	AF210843.1			
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				

JOURNAL	Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE	20130945
PUBMED	10662695
REFERENCE	2 (bases 1 to 68750)
AUTHORS	Molnar, I.
TITLE	Direct Submission
JOURNAL	Submitted (03-DEC-1999) Natural Product Genetics, Novartis Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box 12257, Research Triangle Park, NC 27709, USA
FEATURES	location/Qualifiers
source	1..68750
	/organism="Polysangium cellulosum"
	/strain="So ce90"
	/db_xref="taxon:56"
	<1..1826
CDS	/note="ORF1"
	/codon_start=3
	/transl_table=1
	/product="unknown"
	/protein_id="AAF6904.1"
	/db_xref="GI:6724238"
	/translation="ASIDALFARATSAVLDDGGRATERHVLAKGIEDLRALREH LRIOEGSPSEHCMLGDTVELLADHPDIASISPHNARS.LRHPMTDDAMVDSALVY RWLRAKPGCPRLREYERERERAKTQGRRLMLAATPCRP.PAYMLPENLLGLPLP MSPEVAERLRASVATPELACALLAMSTGGPWSG.PAYMLPENLLGLPLP AIAASAPGTSEALRGAARLFASMEVSSSKSOLGNIPEALMERLTIYAMNADN LSRERERAIYAEVRLRAQAPPAAGAGLAVASSGSLGVTGDLASGDND IVMPGRISPVVLACGDPPEFLAPPLSOMLFVANHAGTISKVTEGSLIYMARNR QARMSYVHARGPMAYNOAMVPDEPGARVYQVRSNIMPEHPTPCIHPPASFS LACDEHLWCESLQSLKLELNRPHHHRGAPSRATYGEHPIATWYPSLTATHTV WADDRRLALGVDRKTGVEPTVLAETHPPAHVSEEDNDIFALGDPDSRDWHETR SCASTVADYQROLDKRPDMVLTNRGCFEFTNDRILTLARS"
	complement(1900..3171)
CDS	/note="ORF2: similar to hypothetical proteins from Mycobacterium tuberculosis and Streptomyces coelicolor and to DD-peptidases"
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAF6905.1"
	/db_xref="GI:6724239"
	/translation="MGALISYAAPGCLAGCAEEGPGQDAGALAPAREVMAEVA AGQDPGAWLVARGDVHDVAVGTLEGGSAIPMRDT1FRIASMTKAVTAAVMLE EGKLDLSDPYDRMLPELANRKVLRIDGPIDETVPABRPIDMLMTFMGFGISFPA SSP1QRIADIDELGVNAQVPWMTPHGPDEMIIRLTGLPLMHQPGQMMYNGSIVQVLA VGRAGDGFPAFVERERILAPICMRDTPDHFVADKLFAPAGCGPTDDEGKTRMDND GAEEAYSPAPFPGGAAGLVSTVDYTLIFARMLNGVHEBERLLISAASVEMRADHL TPAKAASAFEPGFETFHGWGVAVVAATVAVSEVGRYGMWDGGFTSWINDGRELI IGIYMTQASGFLFGALERFWRSYVATESA"
	3415..5556
CDS	/note="ORF3: similar to Na/H and drug/H antiporters"
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAF6906.1"
	/db_xref="GI:6724240"
	/translation="MHGTRQVILSVTLALIVTARASGELARRLROPEVLGELPFG GVYGSVAGLAPGHRALPOBPAVCVSGISGICALLILMAGLEVNDGILRKRA RPGALSLGATAPPLANGAFSALYDRPLDPSGFLDGLVYSVTVASVYAKVLIRESKM RRTAQVYTLAAGVSEVAAMVLAAMTSSSGASPALVARALLASFLGLFWLVGRRR LTHLRKMWADLAFVSSGOVSLVLTFLAALORGLHGLGAFGLVGLINSPATP NRPDLGVQTLVAGLFAPVFLVLAGMDEVSRLQTLPAAMGVALLATAVAAKVPAA IGALVLRGSEALVAVGLMMKGMDLIVAIVVEGLLSNEAYTWAVAAVTVAA SPALLVLEKRAPPTQESARLEPERARAYITGVRRITVYAHALPGAPNDIVSS IVASKRLGETVDTLSEDOAPGSPRAGASRGILARGLRVPVIMRORRLRSS IQALLRASRHDLLVIGARSPPARAGSFGRLDAIVORASENVLVVAGDPAAERAS ARLITVPIGLSEYFAADLAHVALMDLVELSSAQTDGPAVWVRDRSPSVRAV ARVSDAVFGRLEYSRVHAGMSDELITRELARAPYDLVLVIGCDHGLGRLV YLGSTVSVAVRSPVALLVYAHGCTREQYR"
	complement(5612..5992)
CDS	/note="ORF4"
	/codon_start=1
	/transl_table=1

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 7.4e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 60
|||||
Db 33812 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 33871
|||||

QY 61 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 101
|||||
Db 33872 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 33912
|||||

RESULT 4
ARI99551
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 7.4e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 60
|||||
Db 33812 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 33871
|||||

QY 61 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 101
|||||
Db 33872 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 33912
|||||

RESULT 5
ARI99559
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 7.4e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 60
|||||
Db 33812 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 33871
|||||

QY 61 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 101
|||||
Db 33872 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 33912
|||||

|||||
Db 33812 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 33871
|||||

QY 61 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 101
|||||
Db 33872 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 33912
|||||

RESULT 6
ARI99567
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 7.4e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 60
|||||
Db 33812 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 33871
|||||

QY 61 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 101
|||||
Db 33872 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 33912
|||||

RESULT 7
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 7.4e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 60
|||||
Db 33812 GAGACCGAAGCCGATGAGACACCCAGCCCGCGGTGCGAGCCGACAGACGGCGC 33871
|||||

QY 61 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 101
|||||
Db 33872 CCCGGCGCGTTGCCCGCTCGAGTCTGTGAGAGACCGCGC 33912
|||||

Db 33872 CCGGCGCGTTGCCGCGCTCGAGGTGCGAGAGACCGCGC 33912

RESULT 8
AR208671
LOCUS AR208671 68730 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 7,4e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCGCGATGAGACGACCGACGC CGCGGTGCGGTGCGACGACGACGCGCG 60
Db 33812 GAGACCGAAGCGCGATGAGACGACCGACGC CGCGGTGCGGTGCGACGACGACGCGCG 33871
QY 61 CCGGCGCGTTGCCGCGCTCGAGGTGCTGAGGACCGCGC 101
Db 33872 CCGGCGCGTTGCCGCGCTCGAGGTGCTGAGGACCGCGC 33912

RESULT 9
AR172664
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epoethilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1..71989
/organism="unknown"

BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 7,3e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCGCGATGAGACGACCGACGC CGCGGTGCGGTGCGACGACGACGCGCG 60
Db 28200 GAGACCGAAGCGCGATGAGACGACCGACGC CGCGGTGCGGTGCGACGACGACGCGCG 28259
QY 61 CCGGCGCGTTGCCGCGCTCGAGGTGCTGAGGACCGCGC 101
Db 28260 CCGGCGCGTTGCCGCGCTCGAGGTGCTGAGGACCGCGC 28300

RESULT 10
AC119719
LOCUS AC119719 80966 bp DNA linear HTG 23-JUL-2002

DEFINITION Rattus norvegicus clone CH230-525B13, *** SEQUENCING IN PROGRESS
ACCESSION AC119719
VERSION AC119719.4 GI:21902752
KEYWORDS HTG; PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

AUTHORS

1 (bases 1 to 80966)
Muzny,D.M., Adams,C., Adio-Odola,B., Alt-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,X., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvath,J., Kovar,C., Kravtsov,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtenarge,O., Lien,C., Liu,J., Liu,M., Lohlscheid,H., Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maneswarai,M., Mapa,P., Martin,R., Martinale,A., Martinez,E., Massey,E., Mathiney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabati,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okunodu,G., Orangun,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Slason,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Sytek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWTP

```
Center clone name: CH230-525B13
----- Summary Statistics -----
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 31053 bases at least Q40
Consensus quality: 34998 bases at least Q30
Consensus quality: 37207 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.bjsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1029: contig of 1028 bp in length
1128: gap of unknown length
1129: contig of 1035 bp in length
2163: gap of unknown length
2263: contig of 1080 bp in length
3343: gap of unknown length
3443: contig of 1161 bp in length
4604: gap of unknown length
4704: contig of 1579 bp in length
6283: gap of unknown length
6383: contig of 1237 bp in length
7620: gap of unknown length
7721: contig of 1069 bp in length
8790: gap of unknown length
8889: contig of 1322 bp in length
10211: gap of unknown length
10312: contig of 1148 bp in length
11459: gap of unknown length
13008: contig of 1449 bp in length
13108: gap of unknown length
14251: contig of 1143 bp in length
14351: gap of unknown length
15353: contig of 1002 bp in length
15453: gap of unknown length
16525: contig of 1072 bp in length
16625: gap of unknown length
18040: contig of 1415 bp in length
18140: gap of unknown length
19338: contig of 1198 bp in length
19439: gap of unknown length
20604: contig of 1166 bp in length
20704: gap of unknown length
21934: contig of 1230 bp in length
22034: gap of unknown length
23843: contig of 1809 bp in length
23943: gap of unknown length
24958: contig of 1015 bp in length
25058: gap of unknown length
26621: contig of 1563 bp in length
26721: gap of unknown length
28050: contig of 1329 bp in length
28150: gap of unknown length
29585: contig of 1435 bp in length
29685: gap of unknown length
30962: contig of 1277 bp in length
31062: gap of unknown length
32526: contig of 1464 bp in length
32626: gap of unknown length
34036: contig of 1410 bp in length
34136: gap of unknown length
35647: contig of 1511 bp in length
35747: gap of unknown length
37885: contig of 2138 bp in length
37985: gap of unknown length
```

```
*
37986 39418: contig of 1433 bp in length
39419 39518: gap of unknown length
41325 41325: contig of 1807 bp in length
41326 41326: gap of unknown length
41425 41425: contig of 1891 bp in length
43316 43316: gap of unknown length
43317 43317: gap of unknown length
43417 43417: contig of 1941 bp in length
45357 45357: gap of unknown length
45358 45358: gap of unknown length
45458 45458: contig of 1341 bp in length
46798 46798: gap of unknown length
46799 46799: gap of unknown length
46899 46899: contig of 2010 bp in length
48908 48908: gap of unknown length
49009 49009: contig of 1632 bp in length
50640 50640: gap of unknown length
50641 50641: gap of unknown length
50740 50740: contig of 1800 bp in length
52540 52540: gap of unknown length
52640 52640: gap of unknown length
52641 52641: contig of 1471 bp in length
54112 54112: gap of unknown length
54211 54211: contig of 1821 bp in length
56032 56032: gap of unknown length
56132 56132: gap of unknown length
56133 56133: contig of 2512 bp in length
58645 58645: gap of unknown length
58745 58745: contig of 2636 bp in length
61381 61381: gap of unknown length
61481 61481: contig of 3475 bp in length
64955 64955: gap of unknown length
65055 65055: gap of unknown length
67903 67903: contig of 2848 bp in length
67904 67904: gap of unknown length
68003 68003: gap of unknown length
68004 68004: contig of 2434 bp in length
70437 70437: gap of unknown length
70438 70438: contig of 3676 bp in length
70538 70538: gap of unknown length
74213 74213: gap of unknown length
74214 74214: contig of 3165 bp in length
74314 74314: gap of unknown length
77479 77479: gap of unknown length
77578 77578: contig of 3388 bp in length.
77579 77579: Location/Qualifiers
1..80966
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-525B13"

BASE COUNT 22656 a 16980 c 15714 g 21136 t 4480 others
ORIGIN
Query Match 36.4%; Score 36.8; DB 2; Length 80966;
Best Local Similarity 63.6%; Pred. No. 1.3e+02;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Oy 1 GAGACCGAAGCGATGGAGCAGCCAGCCGCGGTGGTGGAGCAGCAGCGGCG 60
Db 17536 GACCAACAATTCCTCGTGAGAGCCGCCCGGAGGGAGGCGACAGCAACCGCGAG 17595

Oy 61 CCGGCGCGGTGCGCGCGCTCGAGGTGC 88
Db 17596 GCCGCGCGCAGCGCGCGCGCGGCGC 17623

RESULT 11
AC105514 155406 bp DNA linear HTG 13-JUL-2002
AC105514/c LOCUS
DEFINITION *** Rattus norvegicus clone CH230-183L2, *** SEQUENCING IN PROGRESS
ACCESSION AC105514.2 GI:21736378
VERSION AC105514.2
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 155406)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooke,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Bouck, J., Bowles, S., Bileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Butay, C., Burch, P., Burkett, C., Burnett, K.L., Byrd, N.C.,
 Cartron, T.F., Carter, M., Carzozo, S.R., Chacko, J., Chavez, C.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dedrich, D.A.H.,
 Delnate, K.R., Delgado, O., Iken, A.L., Ding, Y., Dinh, H.,
 Donawate, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flings, N., Ford, J., Foster, P.,
 Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Giller, R.,
 Correll, J., H. Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haykal, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Huylk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudan, S.,
 Karlsson, E., Kelly, S., Khai, U., King, L., Korvah, J.J., Kovar, C.,
 Krtlovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, H.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, S., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Matheswar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, S.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nookewo, S., Oguh, M., Okunnu, G.,
 Orangun, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Platus, E., Pu, L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojchokhan, I., Rolfe, M., Ruiz, S., Saverly, G.,
 Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Strick, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Teitrod, B., Thomas, N., Thomas, S.,
 Uman, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wlaczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, F., Zhou, J., Zorilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 2 (bases 1 to 155406)
 Worley, K.C.
 Direct Submission
 Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 155406)
 Worley, K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced g1:18092736.

----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GNCG
 Center clone name: CH230-18312
 ----- Summary Statistics -----
 Sequencing vector: Plasmid
 Chemistry: Dye terminator Big Dye: 100% of reads
 Assembly program: Phrap, version 0.990329
 Consensus quality: 105305 bases at least Q40
 Consensus quality: 112375 bases at least Q30
 Consensus quality: 117353 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 69 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

[illegible]

*	5446	5614:	contlg of 1649	bp	in	length
*	5615	5624:	gap of unknown	length		
*	5625	5710:	contlg of 1696	bp	in	length
*	57911	58010:	gap of unknown	length		
*	58011	59278:	contlg of 1268	bp	in	length
*	59279	59378:	gap of unknown	length		
*	59379	61050:	contlg of 1672	bp	in	length
*	61051	61150:	gap of unknown	length		
*	61151	63345:	contlg of 2195	bp	in	length
*	63346	63445:	gap of unknown	length		
*	63446	65331:	contlg of 1986	bp	in	length
*	65331	65531:	gap of unknown	length		
*	65532	67268:	contlg of 1737	bp	in	length
*	67269	67268:	gap of unknown	length		
*	67369	69397:	contlg of 2229	bp	in	length
*	69397	69697:	gap of unknown	length		
*	69698	71840:	contlg of 2143	bp	in	length
*	71841	71940:	gap of unknown	length		
*	71941	73315:	contlg of 1575	bp	in	length
*	73316	73315:	gap of unknown	length		
*	73616	75334:	contlg of 1719	bp	in	length
*	75335	75334:	gap of unknown	length		
*	75435	77060:	contlg of 1626	bp	in	length
*	77061	77160:	gap of unknown	length		
*	77161	79469:	contlg of 2309	bp	in	length
*	79470	79669:	gap of unknown	length		
*	79670	82104:	contlg of 2535	bp	in	length
*	82105	82205:	gap of unknown	length		
*	82205	83687:	contlg of 1483	bp	in	length
*	83688	83787:	gap of unknown	length		
*	83788	83794:	contlg of 3807	bp	in	length
*	83795	87694:	gap of unknown	length		
*	87695	89779:	contlg of 2085	bp	in	length
*	89780	89879:	gap of unknown	length		
*	92763	92763:	contlg of 2884	bp	in	length
*	92864	92863:	gap of unknown	length		
*	92864	95573:	contlg of 2710	bp	in	length
*	95574	95673:	gap of unknown	length		
*	95674	97554:	contlg of 1881	bp	in	length

Query Match	36.2%	Score 36.6;	DB 2;	Length 155+106;
Best Local Similarity	60.6%;	Pred. NO. 1.2e+02;		
Matches 60; Conservative	0;	Mismatches 39;	Indels 0;	Gaps 0;

[illegible]

RESULT 12	
AC128756	
LOCUS	AC128756
DEFINITION	AC128756 92600 bp DNA linear HTG 24-JUL-2002
ACCESSION	Rattus norvegicus clone CH230-187J18, *** SEQUENCING IN PROGRESS
VERSION	AC128756
KEYWORDS	***, 55 unordered pieces.
SOURCE	AC128756.1 GI:21952558
ORGANISM	HTG; HTGS_PHASE1.
	Rattus norvegicus.
	Rattus norvegicus

REFERENCE
AUTHORS
1 (bases 1 to 92600)
Muzny D.M., Adams,C., Adlo-Ogunola,B., Ali-Osman,F.R., Allen,C.,
Aishbrooks,S.L., Ameralungu,H.C., Aye,J.R., Ayele,M., Baks,T.,
Bachajala,J., Benton,J., Bimango,K., Blankenburg,K., Bonini,D.,
Boock,J.A., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buday,R., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Cairton,T.F., Carter,M., Cavacons,S.R., Chacko,J., Chavez,D.,

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 92600)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	----- Genome Center -----

```

Center Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: KAL1
Center clone name: CH230-187J18

----- Summary Statistics -----
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 31018 bases at least Q40
Consensus quality: 33605 bases at least Q30
Consensus quality: 35249 bases at least Q20
-----

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1      1078: contig of 1078 bp in length
*      1079      1178: gap of unknown length
*      1179      2294: contig of 1116 bp in length
*      2295      2394: gap of unknown length
*      3433: contig of 1039 bp in length
*      2395

```

3434 3533: gap of unknown length
3534 4578: contig of 1045 bp in length
4579 4678: gap of unknown length
4679 5846: contig of 1168 bp in length
5847 5946: gap of unknown length
5947 6975: contig of 1029 bp in length
6976 7075: gap of unknown length
7076 8196: contig of 1121 bp in length
8197 8297: gap of unknown length
8298 9377: contig of 1081 bp in length
9378 9477: gap of unknown length
9478 10527: contig of 1050 bp in length
10528 11655: gap of unknown length
11656 11755: contig of 1028 bp in length
11756 12922: gap of unknown length
12923 13022: contig of 1167 bp in length
13023 14141: gap of unknown length
14142 14241: contig of 1119 bp in length
14242 15780: gap of unknown length
15781 15880: contig of 1539 bp in length
15881 17348: gap of unknown length
17349 17448: contig of 1468 bp in length
17449 19154: gap of unknown length
19155 19254: contig of 1706 bp in length
19255 20371: gap of unknown length
20372 20471: contig of 1117 bp in length
20472 21980: gap of unknown length
21981 22080: contig of 1509 bp in length
22081 23513: gap of unknown length
23514 23613: contig of 1433 bp in length
23614 24728: gap of unknown length
24729 24828: contig of 1115 bp in length
24829 26482: gap of unknown length
26483 26582: contig of 1654 bp in length
26583 28587: gap of unknown length
28588 28687: contig of 1985 bp in length
28688 30067: gap of unknown length
30068 30167: contig of 1400 bp in length
30168 31529: gap of unknown length
31529 31629: contig of 1362 bp in length
31630 33120: gap of unknown length
33120 33220: contig of 1491 bp in length
33221 34839: gap of unknown length
34839 34939: contig of 1619 bp in length
34940 36666: gap of unknown length
36667 36766: contig of 1727 bp in length
36767 38267: gap of unknown length
38268 38367: contig of 1501 bp in length
38368 40109: gap of unknown length
40110 40209: contig of 1742 bp in length
40210 41984: gap of unknown length
41985 42084: contig of 1775 bp in length
42085 43687: gap of unknown length
43688 43727: contig of 1543 bp in length
43728 45130: gap of unknown length
45131 45230: contig of 1403 bp in length
45231 46973: gap of unknown length
46974 47073: contig of 1743 bp in length
47074 48323: gap of unknown length
48324 48423: contig of 1250 bp in length
48424 50788: gap of unknown length
50789 50888: contig of 2365 bp in length
50889 53130: gap of unknown length
53131 53230: contig of 2242 bp in length
53231 54419: gap of unknown length
54420 54519: contig of 1189 bp in length
54520 55819: gap of unknown length
55820 55919: contig of 1300 bp in length
55920 57191: gap of unknown length
57192 57291: contig of 1272 bp in length
57292 59542: gap of unknown length
59543 59642: contig of 2251 bp in length

59643 61685: contig of 2043 bp in length
61686 61785: gap of unknown length
61786 64099: contig of 2314 bp in length
64100 64199: gap of unknown length
64200 66339: contig of 2140 bp in length
66340 66439: gap of unknown length
66440 67729: contig of 1290 bp in length
67730 67830: gap of unknown length
67831 69978: contig of 2149 bp in length
69979 70078: gap of unknown length
70079 72142: contig of 2064 bp in length
72143 72243: gap of unknown length
72244 73783: contig of 1541 bp in length
73784 73884: gap of unknown length
73885 75646: contig of 1763 bp in length
75647 75746: gap of unknown length
75747 77543: contig of 1797 bp in length
77544 77644: gap of unknown length
77645 78857: contig of 1214 bp in length
78858 78957: gap of unknown length
78959 80450: contig of 1493 bp in length
80451 80551: gap of unknown length
80552 83143: contig of 2593 bp in length
83144 83243: gap of unknown length
83244 85335: contig of 2092 bp in length
85336 85435: gap of unknown length
85436 87511: contig of 2076 bp in length
87512 87612: gap of unknown length
87613 89843: contig of 2232 bp in length
89844 92600: gap of unknown length
92601 92657: contig of 2657 bp in length

FEATURES
source 1..92600
location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-187J18"

Query Match 35.2%; Score 35.6; DB 2; Length 92600;
Best Local Similarity 62.2%; Pred. No. 2.2e+02;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 6 CGAACCCGATGACGACCGCCGCGTGGTGCAGCCGACGAGCGCGCCGCG 65
Db 868 CGGACGCGGACGACGCGCGCGCGGCGGACGAGACGCGGCGCGCGG 927

OY 66 CGCGTGGCGCGCGCTGAGGTGCTGAGGA 95
Db 928 CTCGACGACCGCCCGCAAGAGCGCGGACGA 957

RESULT 13
AC119661/c
LOCUS AC119661 134366 bp DNA linear HTG 23-JUL-2002
DEFINITION Rattus norvegicus clone CH230-25662, *** SEQUENCING IN PROGRESS
ACCESSION AC119661.5 GI:21902745
VERSION AC119661.5
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 134366)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbarta,J., Benton,J., Blimpage,K., Blenkinsburg,K., Bonini,D.,
Bouck,J., Bowie,S., Bowie,M., Brown,E., Brown,M., Bryant,N.P.,
Buhaq,C., Burck,P., Burkett,C., Burnett,K.L., Byrd,N.C.,
Carson,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Day-Carroll,L., Dederlich,D.A.,

Delaney,R.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,C., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotte,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gadisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Huliyil,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,N., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Lonsdale,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabadi,K., Morgan,H., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Roite,M., Ruiz,S., Savery,G.,
Scheier,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooten,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 134366)
Worley,K.C.

Direct Submission
Submitted (30-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 134366)
Worley,K.C.

Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20467932.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: GMSV
Center Clone name: CH230-256E2
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 83294 bases at least Q40
Consensus quality: 87027 bases at least Q30
Consensus quality: 90347 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1237: contig of 1237 bp in length

1238 1337: gap of unknown length
1338 2424: contig of 1087 bp in length
2425 2524: gap of unknown length
2525 3669: contig of 1145 bp in length
3670 3770: gap of unknown length
3770 4929: contig of 1159 bp in length
4929 5029: gap of unknown length
5029 6170: contig of 1141 bp in length
6170 6270: gap of unknown length
6270 7516: contig of 1246 bp in length
7516 7616: gap of unknown length
7616 8717: contig of 1102 bp in length
8717 8818: gap of unknown length
8818 9941: contig of 1123 bp in length
9941 10040: gap of unknown length
10040 11081: contig of 1041 bp in length
11081 11181: gap of unknown length
11181 12261: contig of 1080 bp in length
12261 12361: gap of unknown length
12361 13506: contig of 1145 bp in length
13506 13606: gap of unknown length
13606 15106: contig of 1500 bp in length
15106 15206: gap of unknown length
15206 16380: contig of 1174 bp in length
16380 16480: gap of unknown length
16480 17966: contig of 1486 bp in length
17966 18066: gap of unknown length
18066 19105: contig of 1039 bp in length
19105 19205: gap of unknown length
19205 20568: contig of 1363 bp in length
20568 21680: gap of unknown length
21680 21780: contig of 1012 bp in length
21780 22947: gap of unknown length
22947 23047: gap of unknown length
23047 24154: contig of 1107 bp in length
24154 24254: gap of unknown length
24254 25444: contig of 1190 bp in length
25444 25544: gap of unknown length
25544 27949: contig of 2405 bp in length
27949 28049: gap of unknown length
28049 29450: contig of 1401 bp in length
29450 30746: gap of unknown length
30746 30846: contig of 1196 bp in length
30846 32179: gap of unknown length
32179 32279: gap of unknown length
32279 33614: contig of 1335 bp in length
33614 33714: gap of unknown length
33714 34960: contig of 1246 bp in length
34960 35060: gap of unknown length
35060 37362: contig of 2302 bp in length
37362 37462: gap of unknown length
37462 39351: contig of 1889 bp in length
39351 39451: gap of unknown length
39451 40558: contig of 1207 bp in length
40558 40758: gap of unknown length
40758 42160: contig of 1402 bp in length
42160 42260: gap of unknown length
42260 44051: contig of 1791 bp in length
44051 44151: gap of unknown length
44151 45525: contig of 1374 bp in length
45525 45625: gap of unknown length
45625 47235: contig of 1630 bp in length
47235 47355: gap of unknown length
47355 49554: contig of 2199 bp in length
49554 49654: gap of unknown length
49654 51956: contig of 2302 bp in length
51956 52056: gap of unknown length
52056 53921: contig of 1865 bp in length
53921 54021: gap of unknown length
54021 55835: contig of 1814 bp in length
55835 55935: gap of unknown length

55936 58618: contig of 2683 bp in length
58618 gap of unknown length
58718 61717: contig of 2999 bp in length
61717 gap of unknown length
61817 63684: contig of 1867 bp in length
63684 gap of unknown length
63785 65949: contig of 2164 bp in length
65949 gap of unknown length
66048 67821: contig of 1773 bp in length
67821 gap of unknown length
67822 70305: contig of 2384 bp in length
70305 gap of unknown length
70306 72600: contig of 2193 bp in length
72600 gap of unknown length
72601 74582: contig of 1882 bp in length
74582 gap of unknown length
74583 78292: contig of 3610 bp in length
78292 gap of unknown length
78293 79935: contig of 1543 bp in length
79935 gap of unknown length
80036 83750: contig of 3715 bp in length
83750 gap of unknown length
83851 87621: contig of 3771 bp in length
87621 gap of unknown length
87722 92168: contig of 4447 bp in length
92168 gap of unknown length
92269 96262: contig of 3994 bp in length
96262 gap of unknown length
96263 99910: contig of 3548 bp in length
99910 gap of unknown length
100010 104092: contig of 4082 bp in length
104092 gap of unknown length
104093 110552: contig of 6360 bp in length

Query Match 34.5%; Score 34.8; DB 2; Length 134366;
Best Local Similarity 62.8%; Pred. No. 2.9e+02;
Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 11 GCCGATGACGACCCAGCCGCGGTGCGAGCGAGCGCGCGCGCGGT 70
DB 49280 GCCGATGACGACCCAGCCGCGGTGCGAGCGAGCGCGCGCGGT 70
QY 71 TGCCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 96
DB 49220 GCG 49195

RESULT 14
AC109839/c 152355 bp DNA linear HTG 17-JUL-2002
LOCUS Rattus norvegicus clone CH210-271M10, *** SEQUENCING IN PROGRESS
DEFINITION *** 55 unordered pieces.

ACCESSION AC109839.3 GI:21743805
VERSION AC109839.3
KEYWORDS HTG: HTGS.PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 152355)
Muzny,D.M., Adams,C., Adio-(duola,B., Ali-osman,F.R., Allen,C.,
Alstrooms,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbetta,J., Benton,J., Blincoe,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,N., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.H.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

TITLE JOURNAL
REFERENCE JOURNAL
AUTHORS JOURNAL
TITLE JOURNAL
AUTHORS JOURNAL

REFERENCE JOURNAL
AUTHORS JOURNAL
TITLE JOURNAL
AUTHORS JOURNAL
COMMENT

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,N., Holloway,C., Hollins,B.,
Hornet,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,L., Kureishi,L., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,T., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,R., Louisged,H.,
Lopez,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maneshwari,M., Manney,B., Martini,R., Martindale,A., Martinez,E.,
Massey,E., Massey,B., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pul,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Wenstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 152355)
Worley,K.C.

Direct Submission
Submitted (08-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 152355)
Worley,K.C.

Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced g1.18846423.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOHA
Center clone name: CH230-271M10
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 11301 bases at least Q40
Consensus quality: 117622 bases at least Q30
Consensus quality: 120952 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1040: contig of 1040 bp in length
* 1041 1140: gap of unknown length
* 1141 2768: contig of 1628 bp in length
* 2769 2868: gap of unknown length

2869 4091: contig of 1223 bp in length
4092 4191: gap of unknown length
4192 5976: contig of 1785 bp in length
5977 6076: gap of unknown length
6077 7167: contig of 1091 bp in length
7168 7268: gap of unknown length
7269 8549: contig of 1282 bp in length
8550 8650: gap of unknown length
8651 10359: contig of 1710 bp in length
10360 10459: gap of unknown length
10460 12098: contig of 1639 bp in length
12099 12199: gap of unknown length
12199 13472: contig of 1274 bp in length
13473 13573: gap of unknown length
13573 15615: contig of 2043 bp in length
15616 15715: gap of unknown length
15716 16923: contig of 1208 bp in length
16924 17023: gap of unknown length
17024 18314: contig of 1291 bp in length
18315 18415: gap of unknown length
18416 19430: contig of 1016 bp in length
19431 19530: gap of unknown length
19531 21598: contig of 2068 bp in length
21599 21698: gap of unknown length
21699 22722: contig of 1024 bp in length
22723 22823: gap of unknown length
22823 24919: contig of 2097 bp in length
24920 25019: gap of unknown length
25020 26524: contig of 1505 bp in length
26525 26624: gap of unknown length
26625 28975: contig of 2351 bp in length
28976 29075: gap of unknown length
29076 30712: contig of 1637 bp in length
30713 30812: gap of unknown length
30813 33147: contig of 2335 bp in length
33148 33247: gap of unknown length
33248 34748: contig of 1501 bp in length
34749 34848: gap of unknown length
34849 36700: contig of 1852 bp in length
36701 36800: gap of unknown length
36801 39406: contig of 2606 bp in length
39407 39506: gap of unknown length
39507 41430: contig of 1924 bp in length
41431 41530: gap of unknown length
41531 43892: contig of 2362 bp in length
43893 43992: gap of unknown length
43993 46871: contig of 2879 bp in length
46872 46971: gap of unknown length
46972 48525: contig of 1554 bp in length
48526 48625: gap of unknown length
48626 51785: contig of 3160 bp in length
51786 51885: gap of unknown length
51886 53510: contig of 1625 bp in length
53511 53610: gap of unknown length
53611 55888: contig of 2278 bp in length
55889 55988: gap of unknown length
55989 58034: contig of 2046 bp in length
58035 58134: gap of unknown length
58135 59998: contig of 1864 bp in length
59999 60098: gap of unknown length
60099 62273: contig of 2175 bp in length
62274 62373: gap of unknown length
62374 64829: contig of 2456 bp in length
64830 64929: gap of unknown length
64930 67734: contig of 2805 bp in length
67735 67834: gap of unknown length
67835 70831: contig of 2997 bp in length
70832 70931: gap of unknown length
70932 73720: contig of 2789 bp in length
73721 73820: gap of unknown length
73821 75955: contig of 2135 bp in length
75956 76055: gap of unknown length
76056 79308: contig of 3253 bp in length

79309 79408: gap of unknown length
79409 82621: contig of 3213 bp in length
82622 82721: gap of unknown length
82722 85703: contig of 2982 bp in length
85704 85803: gap of unknown length
85804 88896: contig of 3093 bp in length
88897 88997: gap of unknown length
88997 92297: contig of 3301 bp in length
92298 92397: gap of unknown length
92398 92398: gap of unknown length
92399 95549: contig of 3152 bp in length
95550 95650: gap of unknown length
95650 99040: contig of 3391 bp in length
99041 99140: gap of unknown length
99141 103191: contig of 4050 bp in length
103191 103290: gap of unknown length
103291 107171: contig of 3881 bp in length
107172 107271: gap of unknown length
107272 111292: contig of 4021 bp in length
111293 111393: gap of unknown length
111393 116175: contig of 4783 bp in length
116176 116275: gap of unknown length
116276 119026: contig of 2751 bp in length
119027 119126: gap of unknown length
124826 124826: contig of 5700 bp in length
124827 124926: gap of unknown length
124927 130259: contig of 5333 bp in length
130260 130359: gap of unknown length
130360 137587: contig of 7228 bp in length
137588 137687: gap of unknown length
137688 141811: contig of 4124 bp in length

Query Match 34.3%; Score 34.6; DB 2; Length 152355;
Best Local Similarity 67.1%; Pred. No. 3e+02;
Matches 49; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 9 ACGCGATGAGCAGACCCGCGCGGTGCGAGCGACAGACGCGCGCGCGC 68
Db 4538 ACGC 4479

Qy 69 GTTCCCGCGCCTC 81
Db 4478 GCAGCGCGCGCGC 4466

RESULT 15
AC008193/c
LOCUS
DEFINITION
02.H.1 mep 94D-94D strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 94 unordered pieces.

ACCESSION
AC008193
VERSION
AC008193.10 GI:19482329
HTG; HTGS_PHASE1.
KEYWORDS
Drosophila melanogaster.
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
Butenhoff, C., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Doyle, C.M., Fartan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Kaira, K.,
Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B.,
Richards, S., Sethi, H., Sylvestras, R.R., Tan, K.H., Webster, D.,
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 154607)
REFERENCE
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Fartan, D.E., Galle, R., George, R.A., Harris, N.L.,

TITLE	JOURNAL	REFERENCE	AUTHORS
Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA (bases 1 to 154607)			Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomutan, M.A., Mazza, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Solt, E., Svrtkas, R.R., Wan, K.H., Weinburg, T., Zheng, R., Zieran, L.L. and Rubin, G.M.
Submitted (15-MAR-2002) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA On Mar 15, 2002 this sequence version replaced gl:6957934. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.flycyc.org/sequence/) or send email to bdgs@fruitfly.berkeley.edu . All contigs in this submission meet the following cutoffs: length >= 200 bases. NOTE: This is a 'working draft' sequence. It currently consists of 94 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.			
404	403: contig of 403 bp in length		
484	483: gap of unknown length		
1061	1060: contig of 577 bp in length		
1141	1140: gap of unknown length		
1704	1703: contig of 563 bp in length		
1784	1783: gap of unknown length		
2223	2222: contig of 439 bp in length		
2303	2302: gap of unknown length		
3070	3069: contig of 767 bp in length		
3150	3149: gap of unknown length		
3578	3577: contig of 428 bp in length		
3658	3657: gap of unknown length		
4407	4406: contig of 749 bp in length		
4487	4486: gap of unknown length		
5645	5644: contig of 1158 bp in length		
5725	5724: gap of unknown length		
6940	6939: contig of 1215 bp in length		
7020	7019: gap of unknown length		
7780	7779: contig of 760 bp in length		
7860	7859: gap of unknown length		
8669	8668: contig of 809 bp in length		
8749	8748: gap of unknown length		
8795	8794: contig of 966 bp in length		
9715	9714: gap of unknown length		
9795	9794: contig of 583 bp in length		
10378	10377: gap of unknown length		
10458	10457: contig of 780 bp in length		
11328	11317: gap of unknown length		
11318	11316: contig of 788 bp in length		
12106	12105: gap of unknown length		
12186	12185: contig of 791 bp in length		
12977	12976: gap of unknown length		
13057	13056: contig of 1266 bp in length		
14323	14322: gap of unknown length		
14402	14401: contig of 802 bp in length		
15205	15204: gap of unknown length		
15285	15284: contig of 1107 bp in length		
16392	16391: gap of unknown length		
16472	16471: contig of 831 bp in length		
17303	17302: gap of unknown length		
17383	17382: contig of 1416 bp in length		
18798	18797: gap of unknown length		
18879	18878: contig of 1310 bp in length		
20189	20188: gap of unknown length		
20269	20268: contig of 1393 bp in length		
21662	21661: gap of unknown length		
21742	21741: contig of 1054 bp in length		
22796	22795: gap of unknown length		
22876	22875: contig of 1047 bp in length		
23923	23922: gap of unknown length		
24003	24002: contig of 1190 bp in length		
25193	25192: gap of unknown length		
25273	25272: contig of 1421 bp in length		
26594	26593: gap of unknown length		
26774	26773: contig of 801 bp in length		
27575	27574: gap of unknown length		
27655	27654: contig of 910 bp in length		
28645	28644: gap of unknown length		
30050	30049: contig of 1405 bp in length		
30130	30129: gap of unknown length		
31676	31675: contig of 1546 bp in length		
31756	31755: gap of unknown length		
33382	33381: contig of 1626 bp in length		
33462	33461: gap of unknown length		
34952	34951: contig of 1490 bp in length		
35032	35031: gap of unknown length		
35164	35163: contig of 1132 bp in length		
36244	36243: gap of unknown length		
38079	38078: contig of 1835 bp in length		
38159	38158: gap of unknown length		
39954	39953: contig of 1795 bp in length		
40034	40033: gap of unknown length		
40983	40982: contig of 949 bp in length		
41063	41062: gap of unknown length		
43013	43012: contig of 1950 bp in length		
43093	43092: gap of unknown length		
44794	44793: contig of 1701 bp in length		
44874	44873: gap of unknown length		
46812	46811: contig of 1838 bp in length		
46812	46811: gap of unknown length		
48612	48611: contig of 1941 bp in length		
48833	48832: gap of unknown length		
50462	50461: contig of 1629 bp in length		
50542	50541: gap of unknown length		
52859	52858: contig of 2317 bp in length		
52939	52938: gap of unknown length		
54832	54831: contig of 1893 bp in length		
54912	54911: gap of unknown length		
56619	56618: contig of 1707 bp in length		
56699	56698: gap of unknown length		
57937	57936: contig of 1158 bp in length		
57937	57936: gap of unknown length		
60796	60795: contig of 2859 bp in length		
60876	60875: gap of unknown length		
62729	62728: contig of 1853 bp in length		
62809	62808: gap of unknown length		
65905	65904: contig of 3096 bp in length		
65985	65984: gap of unknown length		
68417	68416: contig of 2432 bp in length		
68497	68496: gap of unknown length		
72335	72334: contig of 3758 bp in length		
72335	72334: gap of unknown length		
77191	77190: contig of 4776 bp in length		
77191	77190: gap of unknown length		
81642	81641: contig of 4451 bp in length		
81722	81721: gap of unknown length		
85348	85347: contig of 3627 bp in length		
85429	85428: gap of unknown length		
86977	86976: contig of 1548 bp in length		
87057	87056: gap of unknown length		
92459	92458: contig of 5403 bp in length		

* 92460 92539: gap of unknown length
* 92540 99116: contig of 6577 bp in length
* 99117 99196: gap of unknown length
* 99197 104719: contig of 5523 bp in length
* 104720 104799: gap of unknown length
* 104800 111051: contig of 6252 bp in length
* 111052 111131: gap of unknown length
* 111132 114739: contig of 3608 bp in length
* 114740 114819: gap of unknown length
* 114820 117464: contig of 2645 bp in length
* 117465 117544: gap of unknown length
* 117545 125411: contig of 7867 bp in length
* 125412 125491: gap of unknown length
* 125492 132573: contig of 7082 bp in length
* 132574 132653: gap of unknown length
* 132654 133260: contig of 607 bp in length
* 133261 133340: gap of unknown length
* 133341 134073: contig of 733 bp in length
* 134074 134153: gap of unknown length
* 134154 135094: contig of 941 bp in length
* 135095 135174: gap of unknown length
* 135175 135921: contig of 747 bp in length
* 135922 136001: gap of unknown length
* 136002 136693: contig of 692 bp in length
* 136694 136773: gap of unknown length
* 136774 137410: contig of 637 bp in length
* 137411 137490: gap of unknown length

Query Match 34.1%; Score 34.4; DB 2; Length 154607;
Best Local Similarity 60.9%; Pred. No. 3.3e+02;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 10 CGCCGATGAGAGACCCAGCCCGCGTGGGTGCGACCGAGAGAGCGCGCCGCGCGC 69
Db 152836 CGCGGGGGCGCGCGCGCGCGGGGGGGCGCGCGCGCGCGCGCGCGC 152777
OY 70 TTGCCGCGCTCGAGGTGCTGAGAGACCGCGC 101
Db 152776 CGGGCGCGCGCGCGCGCGCGCGCGCGCGC 152745

Search completed: November 5, 2002, 12:33:47
Job time : 717.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_28200_28300

Perfect score: 101

Sequence: 1 gagaccgaacgcgatgac.....gagtgctggagacgcgc 101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued_Patents_MA:*
- 2: /cgn2_6/p/tdata/1/ina/5A_COMB.seq:*
- 3: /cgn2_6/p/tdata/1/ina/5B_COMB.seq:*
- 4: /cgn2_6/p/tdata/1/ina/6A_COMB.seq:*
- 5: /cgn2_6/p/tdata/1/ina/6B_COMB.seq:*
- 6: /cgn2_6/p/tdata/1/ina/PCTUS_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	68750	3 US-09-335-409-1	Sequence 1, Appli
2	101	100.0	68750	4 US-09-568-102-1	Sequence 1, Appli
3	101	100.0	68750	4 US-09-567-969-1	Sequence 1, Appli
4	101	100.0	68750	4 US-09-568-480-1	Sequence 1, Appli
5	101	100.0	68750	4 US-09-568-486-1	Sequence 1, Appli
6	101	100.0	68750	4 US-09-568-472-1	Sequence 1, Appli
7	101	100.0	68750	4 US-09-567-899-1	Sequence 1, Appli
8	101	100.0	71989	4 US-09-443-501A-2	Sequence 1, Appli
9	31	30.7	3381	4 US-09-009-119-1	Sequence 1, Appli
10	31	30.7	3381	4 US-09-371-507-1	Sequence 1, Appli
11	30.4	30.1	285	6 5244792-8	Patent No. 5244792
12	29.8	29.5	23673	4 US-09-773-816-1	Sequence 1, Appli
13	29.8	29.5	33529	4 US-09-144-085-3	Sequence 3, Appli
14	29.6	29.3	1500	4 US-09-593-711A-10	Sequence 10, Appli
15	29	28.7	1015	4 US-09-372-422A-31	Sequence 31, Appli
16	28.6	28.3	2588	2 US-08-796-414B-6	Sequence 6, Appli
17	28.4	28.1	705	4 US-08-998-416-1059	Sequence 1059, Ap
18	28.4	28.1	1701	3 US-09-357-072-1	Sequence 1, Appli
19	28.4	28.1	1701	4 US-08-983-502-1	Sequence 1, Appli
20	28.4	28.1	1701	5 PCT-US95-16542-1	Sequence 1, Appli
21	28.4	28.1	1701	5 PCT-US96-10521-1	Sequence 1, Appli
22	28.4	28.1	1723	4 US-09-199-637A-135	Sequence 135, App
23	28	27.7	612	4 US-09-484-970B-80	Sequence 80, Appl
24	28	27.7	3802	1 US-08-404-354B-2	Sequence 2, Appli
25	28	27.7	3802	1 US-08-314-083B-2	Sequence 2, Appli
26	28	27.7	3802	1 US-08-435-675B-2	Sequence 2, Appli
27	28	27.7	3802	1 US-08-336-257A-4	Sequence 4, Appli

28	28	27.7	3802	3 US-08-884-599-2	Sequence 2, Appli
29	28	27.7	4862	6 5386025-7	Patent No. 5386025
30	28	27.7	4464	2 US-08-400-159-7	Sequence 7, Appli
31	28	27.7	4483	3 US-08-611-729A-7	Sequence 7, Appli
32	27.8	27.5	3383	5 PCT-US95-09098-1	Sequence 1, Appli
33	27.8	27.5	4403765	4 US-09-103-840A-2	Sequence 2, Appli
34	27.8	27.5	4411529	4 US-09-103-840A-1	Sequence 1, Appli
35	27.6	27.3	755	4 US-08-998-416-748	Sequence 748, App
36	27.4	27.1	13842	4 US-09-105-537-30	Sequence 30, Appl
37	27.4	27.1	36778	4 US-09-105-537-5	Sequence 5, Appli
38	27.4	27.1	38506	3 US-09-320-878-19	Sequence 19, Appl
39	27.4	27.1	4403765	4 US-09-103-840A-2	Sequence 2, Appli
40	27.2	26.9	1550	2 US-08-609-443B-17	Sequence 17, Appl
41	27.2	26.9	1550	4 US-08-851-896-17	Sequence 17, Appl
42	27	26.7	554	4 US-09-437-054A-1	Sequence 1, Appli
43	27	26.7	3472	6 5244792-2	Patent No. 5244792
44	27	26.7	5020	1 US-08-459-586-5	Sequence 5, Appli
45	27	26.7	5020	2 US-08-282-696-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
: Sequence 1, Application US/09335409
: Patent No. 6121029
:
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/335, 409
: CURRENT FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-335-409-1

Query Match      100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAGACCGAACCCGATGACGACGACCCCGCGTGGTGACGACGACGACGCGCG 60
Db      33812 GAGACCGAACCCGATGACGACGACCCCGCGTGGTGACGACGACGACGCGCG 33871
QY      61 CCCGCGCGTGGCGCGCGCTCGAGGTGCTGGAGGACCGCGC 101
Db      33872 CCCGCGCGTGGCGCGCGCTCGAGGTGCTGGAGGACCGCGC 33912

RESULT 2
US-09-568-102-1
: Sequence 1, Application US/09568102
: Patent No. 6346404
:
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568, 102
```

CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 60
DB 33812 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 33871

QY 61 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 101
DB 33872 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 33912

RESULT 3
US-09-567-969-1
Sequence 1, Application US/09567969
Patent No. 6355457

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,969
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 60
DB 33812 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 33871

QY 61 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 101
DB 33872 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 33912

RESULT 4
US-09-568-480-1
Sequence 1, Application US/09568480
Patent No. 6355458

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,480
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 60
DB 33812 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 33871

QY 61 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 101
DB 33872 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 33912

RESULT 5
US-09-568-486-1
Sequence 1, Application US/09568486
Patent No. 6355459

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,486
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 60
DB 33812 GAGACCGAAGCGCGATGAGACACCGCCGCGGCGTGGAGCGGAGACGAGCGCGG 33871

QY 61 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 101
DB 33872 CCCGGCGCGTGGCGCGCTCGAGTGTGGAGACCGCGC 33912

RESULT 6
US-09-568-472-1
Sequence 1, Application US/09568472
Patent No. 6358719

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,472
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAACCCGATGAGACACCCAGCCCGCGTGGTGACACCGACAGAGCGCGC 60
Db 33812 GAGACCGAACCCGATGAGACACCCAGCCCGCGTGGTGACACCGACAGAGCGCGC 33871

QY 61 CCGGCGCGTGGCGCGCTCGAGGTGCTGGAGAGCGCGC 101
Db 33872 CCGGCGCGTGGCGCGCTCGAGGTGCTGGAGAGCGCGC 33912

RESULT 7
US-09-567-899-1
Sequence 1, Application US/09567899
Patent No. 6383787
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,899
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAACCCGATGAGACACCCAGCCCGCGTGGTGACACCGACAGAGCGCGC 60
Db 33812 GAGACCGAACCCGATGAGACACCCAGCCCGCGTGGTGACACCGACAGAGCGCGC 33871

QY 61 CCGGCGCGTGGCGCGCTCGAGGTGCTGGAGAGCGCGC 101
Db 33872 CCGGCGCGTGGCGCGCTCGAGGTGCTGGAGAGCGCGC 33912

RESULT 8
US-09-443-501A-2
Sequence 2, Application US/09443501A
Patent No. 6303342
GENERAL INFORMATION:

APPLICANT: Kosan Biosciences, Inc.
APPLICANT: Julien, Bryan
APPLICANT: Katz, Leonard
APPLICANT: Khosla, Chaitan
APPLICANT: Tang, Li
APPLICANT: Ziermann, Rainer
TITLE OF INVENTION: Recombinant Methods and Materials for Producing
FILE REFERENCE: 30062-20031.00
CURRENT APPLICATION NUMBER: US/09/443,501A
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: US 60/130,560
PRIOR FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: US 60/122,620
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/119,386
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: US 60/109,401
PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 71989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 4.2e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGACCGAACCCGATGAGACACCCAGCCCGCGTGGTGACACCGACAGAGCGCGC 60
Db 28200 GAGACCGAACCCGATGAGACACCCAGCCCGCGTGGTGACACCGACAGAGCGCGC 28259

QY 61 CCGGCGCGTGGCGCGCTCGAGGTGCTGGAGAGCGCGC 101
Db 28260 CCGGCGCGTGGCGCGCTCGAGGTGCTGGAGAGCGCGC 28300

RESULT 9
US-09-009-119-1
Sequence 1, Application US/09009119
Patent No. 6160206
GENERAL INFORMATION:
APPLICANT: SATO, Ryo
APPLICANT: BOYNTON, John E.
APPLICANT: GILHAM, Nicholas W.
APPLICANT: HARRIS, Elizabeth H.
TITLE OF INVENTION: Porphyria Accumulating-Type Herbicide Resistance Gene
FILE REFERENCE: substitute sequence listing
CURRENT APPLICATION NUMBER: US/09/009,119
CURRENT FILING DATE: 1998-01-20
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 3381
TYPE: DNA
ORGANISM: Chlamydomonas reinhardtii
US-09-009-119-1

Query Match 30.7%; Score 31; DB 4; Length 3381;
Best Local Similarity 62.0%; Pred. No. 3.5;
Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 23 CCGACGCGCGGTGGGTGACCGGACAGAGCGCGCGCGCGCTTGC 82
Db 2982 CCGACGCGGTGGGTGACCGGACAGAGCGCGCGCGCGCTTGC 3041
QY 83 AGGTGCTGAGAGCGCGC 101
| | | | |

[illegible]

Db 18311 GGACGCGCGCGCGCGAGGTGGGCTGTACAGC 18347

RESULT 14

US-09-593-711A-10/c
; Sequence 10, Application US/09593711A
; Patent No. 6271030

GENERAL INFORMATION:

; APPLICANT: Brett P. Moria
; APPLICANT: Madeline M. Butler
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF C/EBP BETA EXPRESSION
; FILE REFERENCE: RTS-0118
; CURRENT APPLICATION NUMBER: US/09/593,711A
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 244
; SEQ ID NO 10
; LENGTH: 1500
; TYPE: DNA
; ORGANISM: Mus musculus

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (108)...(998)
US-09-593-711A-10

Query Match

29.3%; Score 29.6; DB 4; Length 1500;
Best Local Similarity 61.8%; Pred. No. 8.1;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 16 TGAGAGACCCAGCCGCGGTGCGACCCGCGCGCGCGGTGCGG 75

DB 739 TTGGGGGGGCGCGCGCGCGCGCGAGAGAGCGCGCGCGCTTGGCGTGGCG 680

QY 76 CGCCTCGAGGTGCTG 91

DB 679 GGGCTCGCGCTGCCG 664

RESULT 15

US-09-372-422A-31/c
; Sequence 31, Application US/09372422A
; Patent No. 6313375

GENERAL INFORMATION:

; APPLICANT: Rudolf Jung
; APPLICANT: Francois Barrien
; TITLE OF INVENTION: Maize Aquaporins and Uses Thereof
; FILE REFERENCE: 0919
; CURRENT APPLICATION NUMBER: US/09/372,422A
; CURRENT FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: US 60/098,692
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 1015
; TYPE: DNA
; ORGANISM: Zea mays

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (77)...(863)
US-09-372-422A-31

Query Match

28.7%; Score 29; DB 4; Length 1015;
Best Local Similarity 55.4%; Pred. No. 12;
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

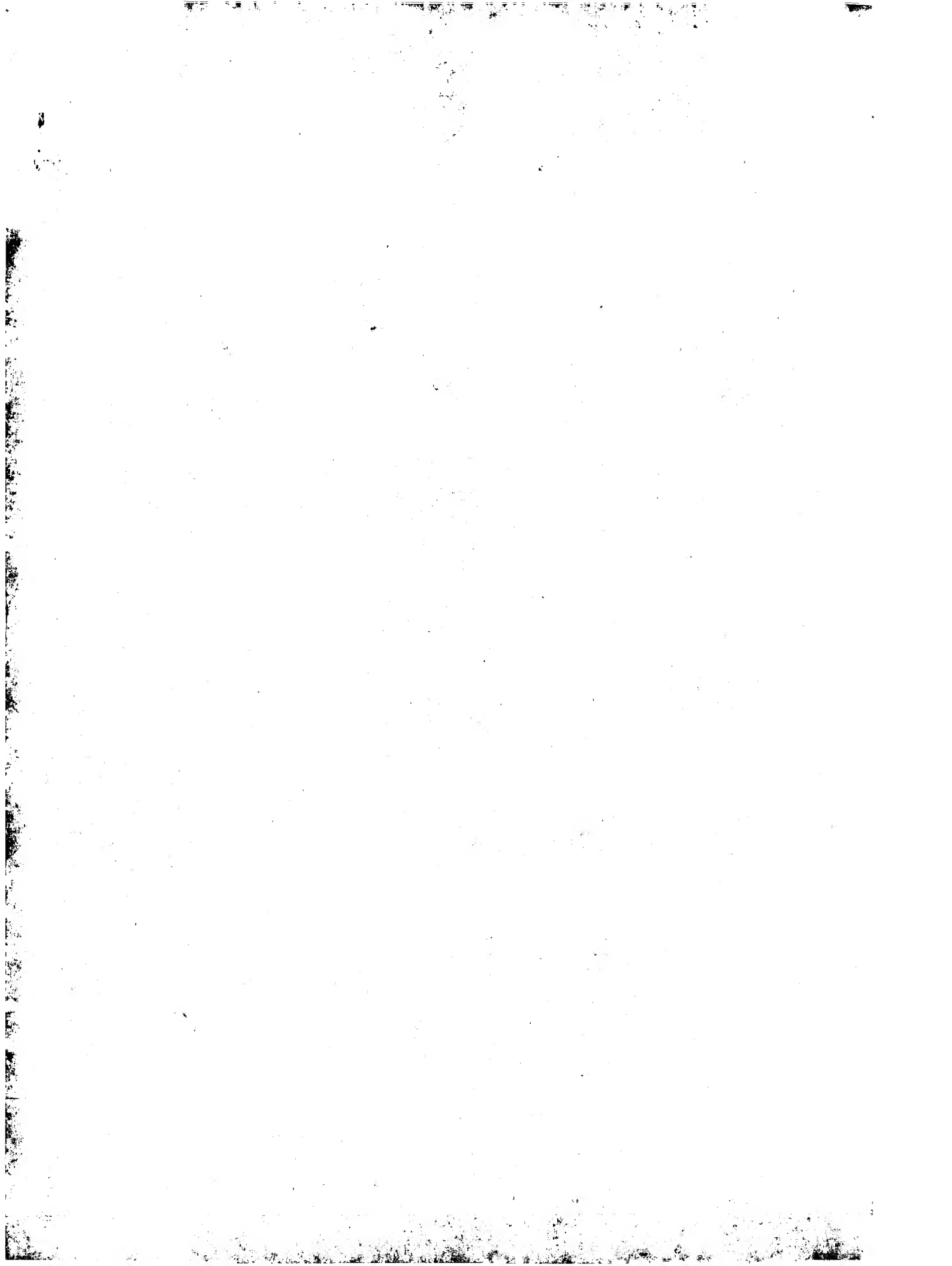
QY 1 GAGACCGAGCGCGAGTGGAGACCAAGCCGCGGTGCGCGCGAGCCGAGACGCGCG 60

DB 493 GAGCGGAACCCGCGGCGCATGCCCCGCGGAGAGCGAGCGAGCGTGGCGGC 434

QY 61 CCGGCGCGTTCGCGCGCTCGAGGTGCTGAGAGACCGCGC 101

Db 433 GACGGCGCCAGACAGCTGCGCGAGCCAGTACAGACCGCGC 393

Search completed: November 5, 2002, 15:43:44
JOB Time : 147.737 secs




```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3608765"
/clone_1id="NH MRC 44"
/tissue_type="endometrium, endocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notice="Organ: uterus; Vector: pORF; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

```

Query Match	38.4%	Score	38.8	DB	10	length	639
Best Local Similarity	64.4%	Pred.	No. 9.5				
Matches	58	Conservative	0	Mismatches	32	Indels	0
						Gaps	0
Oy	11	GGCGATGGACGAGCCACAGCGCCGCGGTGCGTCGTACACCGACAGACGCGCGCGCGCGCT	70				
Db	185	GGCAGAGGACGACCTGCGGCTCCGCCCGCTGTCGCCGACAGACCGCAGCGCGGACG	244				
Oy	71	TGCGCGGCTCTCGAGTGTGCTGAGAGACCGTG	100				
Db	245	AGGACCGCGCGGCGGCTGCGAGGTGAGCTGG	274				

RESULT 2	978 bp	mrna	linear	EST 12-JAN-2001
BF795540/c				
LOCUS				
DEFINITION	602255474e1 NIH_MGC_85 Homo sapiens CDNA clone			IMAGE:4342603 5',
ACCESSION	BF795540			
VERSION	BF795540			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 978)			
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .			
TITLE	National Institutes of Health. Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Louis Staudt, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLN) Clone sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLN at: http://image.lln.gov Plate: L14M9398 row: c column: 20 High quality sequence stop: 687. Location/Qualifiers:			
FEATURES				

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4342403"
/clone_id="NIM_MGC_85"
/tissue_type="lymphoma, cell_line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Oligo-5' primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies
Note: this is a NIM_MGC library."

```

Best Locally Similarity 61.4%; Pred. No. 9.7; Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY	1	GAGACCGAAGCGCGGATGACGACCCACGCGCGGTCTGGGTCTCAGCCGACAGACGCGCGG	60
Db	975	GAGCGCACCCCGGGAACAGAGGACCCCGCGCCGCGGACCGGACCGACCGCGCGCTC	916
QY	61	CCCGCGCGCTTTCGCCGCTCTGAGGTGCTGAGACCGCGCC	101
Db	915	CGGCGCGCGCGCTTCCCTGTGACGACCGGCTGGGCGCCC	875

RESULT	3
AG030608	
LOCUS	1189 bp DNA linear GSS 01-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-003A18.R, genomic survey sequence.
ACCESSION	AG030608
VERSION	AG030608.1 GI:16557481
KEYWORDS	GSS:
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male BAC Library Clone:PTB-003A18.R,
ORGANISM	Pan troglodytes

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoiki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of library PTB
Unpublished
2 (bases 1 to 1189)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoiki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

1-7-22 suenaka-enou,tsukuma-ku, tokioka, natsugawa, riken.go.jp, E-mail:chlimpb@gsc.riken.go.jp, URL:http://npg.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PNG This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.
 PRIMERS

```

sequencing: mixrev
LIBRARY
  Vector      : pRS145
  R.Site 1    : SacI
  R.Site 2    : SacI.
Location/Qualifiers
  1. .1189
  /organism="Pan troglodytes"
  /db_xref="taxon:9598"
  /clone="pmb-003A18.R"
  /sex="male"
  /cell_type="lymphoblast"
  /clone_1lb="Pfb Chimpanzee Male BAC Library"
BASE COUNT      17 a      488 c      540 g      24 t      120 others
ORIGIN

```

[illegible]

RESULT 4
BG809572

LOCUS BG09572 888 bp mRNA linear EST 22-MAY-2001
DEFINITION mgct001xal4f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
ACCESSION BG09572
KEYWORDS BG09572.1 GI:14180552
SOURCE EST.
ORGANISM Magnaporthe grisea.
Magnaporthe grisea.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes; incertae sedis; Magnaportheaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 888)
AUTHORS Choi,W. and Dean,R.A.
TITLE Construction and sequence analysis of an appressorium stage cDNA
JOURNAL library in the rice blast fungus, Magnaporthe grisea
COMMENT Unpublished (2001)
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email:ralph.dean@ncsu.edu
Seq primer: T3 primer (AATTAACTCTCAAGG).
FEATURES
source
1..888
/organism="Magnaporthe grisea"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgct001xal4f"
/clone_id="Magnaporthe grisea Appressorium Stage cDNA"
/dev_stage="Germinated conidia on appressorium-inductive
surface"
/note="Vector: pBluescript SK(+) Vector; Site_1: EORI;
Site_2: XhoI; The appressorium formation specific cDNA
library was constructed from conidia germinated for 5-8 hr
on an inductive surface. The library contains over 55,000
clones with average insert size of 1.5 kbp."
BASE COUNT 22 a 431 c 397 g 4 t 34 others
ORIGIN
Query Match 35.8%; Score 36.2; DB 12; Length 888;
Best Local Similarity 60.8%; Pred. No. 35;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
OY 5 CCGAAGCGGAGTGCAGACCCAGCGCGGTCGACCGACACAGCGCGCGCG 64
DB 623 CCGGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 682
OY 65 GCGCGTTCGCGCGCTCGAGGTGCTGAGAGACCGCGC 101
DB 683 CCGC 719

TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbbs@sc.riken.go.jp, URL:http://npg.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : Saci
R.Site 2 : Saci
FEATURES
source
1..813
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-007B17.F"
/sex="male"
/cell_type="lymphoblast"
/clone_id="PTB Chimpanzee Male BAC Library"
BASE COUNT 33 a 320 c 267 g 164 t 29 others
ORIGIN
Query Match 34.7%; Score 35; DB 17; Length 813;
Best Local Similarity 59.6%; Pred. No. 66;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY 12 CCGATGAGACGACCCAGCGCGCGGTGTCGACCGACAGACGCGCGCGCGCTT 71
DB 334 CCGGAGGCG 275
OY 72 GCGCGCGCTCGAGGTGCTGAGAGACCGCG 100
DB 274 GGNCG 246

RESULT 6
LOCUS AG063051/c 1020 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-051J10.F, genomic survey sequence.
ACCESSION AG063051
VERSION AG063051.1 GI:16614853
KEYWORDS GSS.
ORGANISM Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC library clone:PTB-051J10.F.
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
COMMENT 2 (bases 1 to 1020)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbbs@sc.riken.go.jp, URL:http://npg.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : Saci

FEATURES R.Site 2 : Sac1
Location/Qualifiers
source 1. 1020
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-051010.F"
/sex="male"
/cell_type="lymphoblast"
/clone_1lb="PTB Chimpanzee Male BAC Library"
BASE COUNT 57 a 377 c 385 g 44 t 157 others
ORIGIN

Query Match 34.7%; Score 35; DB 17; Length 1020;
Best Local Similarity 60.9%; Pred. No. 63;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 5 CCGAACCCGATGAGACACCCCGCGCGTGGTGACCGACAGAGCGCGCGCG 64
DB 985 CCG 926
QY 65 GCGCGTTCG 96
DB 925 GCCCG 894

RESULT 7
LOCUS BO680479 1053 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOUNT_8209830 NIH_MGC_112 Homo sapiens cDNA IMAGE:6272780
ACCESSION BO680479
VERSION BO680479.1 GI:21793158
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1059)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTI
CDNA Library Preparation: Rubin Laboratory
CDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM2451 row: k column: 21
High quality sequence stop: 286.
Location/Qualifiers:

FEATURES
source 1. 1059
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6272780"
/clone_1lb="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOT7; Site:1. XhoI; Site:2.
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

BASE COUNT 88 a 194 c 311 g 78 t 388 others
ORIGIN
Query Match 34.3%; Score 34.6; DB 14; Length 1059;
Best Local Similarity 64.2%; Pred. No. 77;

Matches 52; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1 GAGACCGAACCCCGGATGAGACACCCCGCGCGTGTGGTGACCGACAGACCGCGCG 60
DB 852 GCG 911
QY 61 CCGCGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
DB 912 CCG 932

RESULT 8
LOCUS AV618417 260 bp mRNA linear EST 28-NOV-2001
DEFINITION AV618417 Bos taurus rumen fetus Bos taurus cDNA E08U001G05
5', mRNA sequence.
ACCESSION AV618417
VERSION AV618417.1 GI:9754087
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 260)
REFERENCE Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Ago,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
21570554

JOURNAL
MEDLINE
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ccoca.ocn.ne.jp
Single pass sequencing.
Location/Qualifiers

FEATURES
source 1. 260
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="E08U001G05"
/clone_1lb="Bos taurus rumen fetus"
/tissue_type="rumen"
/dev_stage="fetus"
/lab_host="DH10B"
/note="Vector: pZL1; Site:1: SalI; Site:2: NotI"

BASE COUNT 43 a 99 c 86 g 27 t 5 others
ORIGIN
Query Match 34.1%; Score 34.4; DB 10; Length 260;
Best Local Similarity 67.1%; Pred. No. 11e+02;
Matches 47; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 24 CCGACCG 83
DB 19 CCG 78
QY 84 GGTGCTGAG 93
DB 79 GGGCGCGGAG 88

RESULT 9
LOCUS AG082451 983 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-079F14.F, genomic survey sequence.
ACCESSION AG082451
VERSION AG082451.1 GI:16634253
KEYWORDS GSS.

SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
BAC library clone:PTB-079F14.F.
ORGANISM Pan troglodytes
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.
AUTHORS 1
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Toki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 983)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Toki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
source Location/Qualifiers
1..983
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-079F14.F"
/sex="male"
/cell_type="lymphoblast"
/clone_11b="PTB Chimpanzee Male BAC library"
BASE COUNT 128 a 391 c 439 g 15 t 10 others
ORIGIN
Query Match 33.7%; Score 34; DB 17; Length 983;
Best Local Similarity 63.4%; Pred. No. 1.1e+02;
Matches 52; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 6 CGAAGCGGATGAGCGACCGCGGGTGCAGCGAGCGCGCGCCGG 65
Db 289 CGAAGCGGAGGAGCGACCGCGCGCGCGCGCGCGCGCGAGCGCGCGCG 348
QY 66 CGCGTTGCCGCGCTCGAGGTG 87
Db 349 GCGCGGAGGAGCGCGCGCGCG 370
RESULT 10
AJ449574 378 bp mRNA linear EST 19-APR-2002
LOCUS AJ449574 riken1 Gallus gallus cdna clone 22h1rl, mRNA sequence.
DEFINITION AJ449574
ACCESSION AJ449574
VERSION AJ449574.1 GI:20216795
KEYWORDS EST.
SOURCE chicken
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 378)
AUTHORS Buerstedde,J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)
CONTACT: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

FEATURES
source Location/Qualifiers
1..378
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="22h1rl"
/clone_11b="riken1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 57 a 136 c 140 g 45 t
ORIGIN
Query Match 33.5%; Score 33.8; DB 9; Length 378;
Best Local Similarity 62.4%; Pred. No. 1.4e+02;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 17 GGAGACCCAGCCCGCGGTGCGAGCGAGAGCGCGCGCGCGTTCGCCG 76
Db 93 GGCTGACCGAGCCAAACCCGCGAGACGTGCGACGCCGCGCTGAGCGAGCTCC 152
QY 77 GCCTGAGGTGCTGAGAGCCGCC 101
Db 153 GCCAGCCGTGCTGACAGCCGCC 177
RESULT 11
AJ452432 681 bp mRNA linear EST 22-APR-2002
LOCUS AJ452432 riken1 Gallus gallus cdna clone 3006rl, mRNA sequence.
DEFINITION AJ452432
ACCESSION AJ452432
VERSION AJ452432.1 GI:20262528
KEYWORDS EST.
SOURCE chicken
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 681)
AUTHORS Buerstedde,J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)
CONTACT: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
source Location/Qualifiers
1..681
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="3006rl"
/clone_11b="riken1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 131 a 198 c 262 g 89 t 1 others
ORIGIN
Query Match 33.5%; Score 33.8; DB 9; Length 681;
Best Local Similarity 62.4%; Pred. No. 1.3e+02;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 17 GGAGACCCAGCCCGCGGTGCGAGCGAGAGCGCGCGCGCGTTCGCCG 76
Db 93 GGCTGACCGAGCCAAACCCGCGAGACGTGCGACGCCGCGCTGAGCGAGCTCC 152
QY 77 GCCTGAGGTGCTGAGAGCCGCC 101
Db 153 GCCAGCCGTGCTGACAGCCGCC 177
RESULT 12
AQ272425/c

LOCUS AC272425 714 bp DNA linear GSS 03-NOV-1996
DEFINITION nbxb0027p16f CUG1 Rice BAC library Oryza sativa genomic clone
ACCESSION nbxb0027p16f, DNA sequence
AC272425
AC272425.1 GI:3825740
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Etreptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidoideae; Oryzaceae; Oryza.
1 (bases 1 to 724)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGG
Class: BAC ends
High quality sequence stop: 136.
Location/Qualifiers
1..724
FEATURES
source

12 a 287 c 300 g 95 t 30 others

Query Match	33.3%	Score 33.6;	DB 17;	Length 724;
Best Local Similarity	59.3%	Pred. NO. 1.4e+02;		
Matches 54;	Conservative 0;	Mismatches 37;	Indels 0;	Gaps 0;
QY	11	GCCGATGAGACAGCCACGCGCCGGCTGGGATGCACCCACAGACGCGCGCCGCGCGCT	70	
Db	418	GCCCCGGGGGCGCCCCCGCGCNCGGCCGCCCGCCGCGCGCGCGCNCGCCGCGGGGGCGC	359	
QY	71	TGCCCGCGCCCTCGAGGTCCTGGAGGACCGCGC	101	
Db	358	GCGCGCGCGCGCCCGC	328	

RESULT 13
AG137902
LOCUS

AG137902 813 bp DNA linear GSS 04-NOV-2001

DEFINITION	ORGANISM
Pan troglodytes DNA, clone: PTB-151019.R, genomic survey sequence.	
ACCESSION	
AG137902	
VERSION	
AG137902.1	GI:16667580
KEYWORDS	
SOURCE	
	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
	BAC library clone:PTB-151019.R.
	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE	
AUTHORS	1
	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
TITLE	Toyoki, Y., Watanabe, H. and Sakaki, Y.
JOURNAL	BAC end sequences of library PTB
REFERENCE	unpublished
AUTHORS	2 (bases 1 to 813)
	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
TITLE	Toyoki, Y., Watanabe, H. and Sakaki, Y.
JOURNAL	Direct Submission
	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
	1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
	(E-mail: shimpbes@sc.riken.go.jp, URL: http://hnp.gsc.riken.go.jp/ ,
	Tel:81-43-503-9111, Fax:81-43-503-9170)
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end
	was generated during the R&D process and may have higher chance of
	clone tracking errors.
	PRIMERS

```

Sequencing: M3Rev
LIBRARY
Vector      : PKS145
R.Site 1    : SacI
R.Site 2    : SacI.
Location/Qualifiers
1. .813
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-151019.R"
/sex="male"
/cell_type="Lymphoblast"
/clone_idb="PTB Chimpanzee Male BAC Library"
BASE COUNT   174 a     205 c     303 g     20 t     111 others
ORIGIN

Query Match          33.3% Score 33.6; DB 17; Length 813;
Best Local Similarity 56.4%; Pred. No.1.4e+02;
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY      1 GAGACCGAACCCGATGAGCAGCACCCACGCCCGGTGGTGTCAGCCGACAGACGC GGCG 60
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       399 GNGACGGCNCNCNCGGGGGGGCCCCCGGNNCCGGGGCGCCCGGAGACCCC CGCGCGCGC 458

QY      61 CCGGCGCGTGGCCGCGCGCTCGAGGTGCTGGAGGACCGCGC 101
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       459 CCGGCGCGGGGGCGCCGCCCGCCCGCGCGCAAGCGCGCGC 499

```

RESIDUAL	14
CNS006ON	Locus
DEFINITION	CNS006ON 910 bp DNA linear GSS 03-JUN-1999 Drosophila melanogaster genome survey sequence T7 end of BAC # BACR1J21 of RPOI-38 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL065629
VERSION	AL065629.1 GI:4944698
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphytoidea; Drosophilidae; Drosophila. 1 (bases 1 to 910) Genoscope. Direct Submission
REFERENCE	
AUTHORS	
TITLE	

JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqef@genoscope.cns.fr - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

```
1..910
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACR14J21"
/clone_lib="RPCI-98"
/note="end : 77"
```

BASE COUNT 202 a 63 c 112 g 198 t 335 others

ORIGIN

Query Match 33.3%; Score 33.6; DB 17; Length 910;
Best Local Similarity 28.6%; Pred. No. 1.3e+02;
Matches 20; Conservative 37; Mismatches 13; Indels 0; Gaps 0;

OY 32 GGGGTCGGTGCAGCCGAGACGCGCGCCCGGGCGCTGCGCGCGCTCGAGGTCGTG 91

DB 538 GTGGTGGGGGGGSSSSSSSSSSSSCGSSBGBSSGCGSGCTGAGTGTG 597

OY 92 AGGACCGCGC 101

DB 598 CGCGSSSSSS 607

RESULT 15 643 bp mRNA linear EST 26-JUL-2002
BQ788818/C
LOCUS BQ788818
DEFINITION WHE4154.B12.D2425 wheat CS whole plant cDNA library Triticum
aestivum CDNA clone WHE4154.B12.D24, mRNA sequence.
ACCESSION BQ788818.1 GI:21997290
VERSION BQ788818.1
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

REFERENCE
AUTHORS Anderson O.D., Akhunov E., Chao S., Crossman C., Deal K., Dvorak J.,
Lazo G.R., Rausch C.J., Wilson C. and Woo J.
TITLE The structure and function of the expressed portion of the wheat
JOURNAL genomes - Chinese Spring whole plant cDNA library
COMMENT unpublished (2002)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 510595773
Fax: 510595818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.
Location/Qualifiers
1..643
/organism="Triticum aestivum"

FEATURES

source

```
1..643
/organism="Triticum aestivum"
```

```
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4154.B12.D24"
/clone_lib="Wheat CS whole plant cDNA library"
/tissue_type="Roots, leaves, crown, stem and sheath"
/dev_stage="Adult"
/lab_host="E. coli SOLR"
/note="Vector: lambda Uni-ZAP XR, excised phagemid  
pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; plant  
tissues from wheat cv. CS grown to full tillering stage in  
greenhouse were collected at University of California,  
Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total  
RNA was prepared from leaves (young leaf and third leaf),  
whole roots, crown, stem and sheath tissues, and then  
equal quantities of RNA were pooled from these  
samples. PolyA was purified from the pooled RNA, a cDNA  
library was made, and the cDNA clones were in vivo excised  
to give pBluescript SK(-) phagemids in J. Dvorak's lab (E.  
Akhunov, J. Dvorak) at the University of California,  
Davis. Colony plating, plasmid DNA preparations and DNA  
sequencing were performed in the OD Anderson lab (all  
other authors)."
```

BASE COUNT 120 a 211 c 218 g 94 t

ORIGIN

Query Match 32.9%; Score 33.2; DB 14; Length 643;
Best Local Similarity 61.6%; Pred. No. 1.8e+02;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 5 CCGACGCGGATGAGACGCCAGCCCGCGGTCGCGAGCGGACGCGCGCGCG 64

DB 207 CCGACGCGGAGAGACGCCAGCGCGCGGCGGAGATCGCGAGACCGCGCGCGCTCG 148

OY 65 GCGCGTGGCGGCGCTCGAGGTCGTG 90

DB 147 CCGCGAGCCCTGTCTGGCGATGCTG 122

Search completed: November 5, 2002, 16:18:18
Job time : 1118.44 secs

[illegible]

```
FT      /product= "Orf 5 protein (AAV58584)"
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      CDS
FT      11872..116104
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      CDS
FT      16251..21745
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      CDS
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      CDS
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      CDS
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      CDS
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      CDS
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      CDS
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      CDS
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      CDS
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      CDS
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
FT
FT
XX      W09966028-A2.
XX
XX      23-DEC-1999.
XX
XX      16-JUN-1999; 99WO-EP04171.
XX
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
XX      05-FEB-1999; 99US-0118906.
XX
XX      (NOVS ) NOVARTIS AG.
```

```
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERM GES MBH.
XX
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D:
PI
XX      WPI: 2000-097741/08.
XX      DR      AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
XX      P-PSDB; AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
XX      AAV58585, AAV58586, AAV58587, AAV58588, AAV58589, AAV58591,
XX      AAV58592, AAV58593, AAV58594.
XX
XX      New isolated epothonone synthase genes, used for the recombinant
PT      production of epothonone for use in cancer therapy.
XX
XX      Claim 14; Page 87-104; 174pp; English.
XX
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC      comprising 22 open reading frames (ORFs) and includes genes encoding
CC      proteins involved in the biosynthesis of epothonones. Epothonones A and
CC      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC      starter unit; polyketides being synthesised from two-carbon building
CC      blocks, the beta-carbon of which always carries a keto group. Each round
CC      of two-carbon addition is carried out by a complex of enzymes known as
CC      the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
CC      the thiazole ring formation of epothonones, and EPOS B, EPOS C, EPOS D
CC      and EPOS E (AAV58575-Y58578) are involved in polyketide backbone
CC      formation. EPO F (AAV58579) is an epothonone macroketone oxidase, and
CC      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
CC      involved in transport. Epothonones mimic the biological activity of
CC      taxol, and may be substituted for taxol in cancer chemotherapeutic
CC      compositions. Epothonones exhibit a much lower drop in potency against a
CC      multiply drug-resistant cell line compared with taxol, and are
CC      considerably less efficiently exported from such cells by the multidrug
CC      resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC      epothonones as anticancer agents, they are problematical to produce on a
CC      large scale. Epothonones are too complex for industrial scale chemical
CC      synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC      poor yields of epothonones. The nucleic acids of the invention may be
CC      used for the recombinant production of epothonones in a heterologous host
CC      that is more amenable to fermentation.
XX
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other:
SQ
XX
XX      Query Match 100.0%; Score 101; DB 21; Length 68750;
XX      Best Local Similarity 100.0%; Pred. No. 1.2e-16;
XX      Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      1 GAGACCGAAGCGGATGAGACGACCCAGCCCGCGGTGCGAGCCGACGACGCGCGG 60
XX      |||||||
XX      DB 33812 GAGACCGAAGCGGATGAGACGACCCAGCCCGCGGTGCGAGCCGACGACGCGCGG 33871
XX      |||||||
XX      QY      61 CCCGGCGGCTTGCCTGCGCGCTCGAGTGTGAGAGACGCGGC 101
XX      |||||||
XX      DB 33872 CCCGGCGGCTTGCCTGCGCGCTCGAGTGTGAGAGACGCGGC 33912
XX
XX      RESULT 2
XX      ID      AAA29349 standard; DNA; 71989 BP.
XX
XX      AC      AAA29349;
XX
XX      DT      12-SEP-2000 (first entry)
XX
XX      DE      Sorangium cellulosum epothonone polyketide synthase operon genomic DNA.
XX
XX      KW      Epothonone; polyketide synthase; epoa; epob; epoc; epod; epoe; epof;
XX      epol; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
XX      tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
XX      OS      Sorangium cellulosum.
XX
XX      FH      Key Location/Qualifiers
```

```
FT CDS 3..992
FT /tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /tag= c
FT /label= epOA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5810
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6135
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493
FT /tag= h
FT /label= epOB_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137
FT /tag= x
FT /label= epOC_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epOD_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574
FT /tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287
FT /tag= ap
```

```
FT      /note= "encodes ER5"
FT      30321..30869
FT      /*tag= aq
FT      /note= "encodes KR5"
FT      31077..31373
FT      /*tag= ar
FT      /note= "encodes A2p5"
FT      31440..32807
FT      /*tag= as
FT      /note= "encodes K36"
FT      33018..34067
FT      /*tag= at
FT      /note= "encodes A16"
FT      34107..34676
FT      /*tag= au
FT      /note= "encodes D116"
FT      35760..36641
FT      /*tag= av
FT      /note= "encodes E16"
FT      36705..37256
FT      /*tag= aw
FT      /note= "encodes K16"
FT      37470..37769
FT      /*tag= ax
FT      /note= "encodes ACp6"
FT      37912..49308
FT      /*tag= ay
FT      /label= epof_gene
FT      /note= "encodes mdclules 7 and 8"
FT      38014..39375
FT      /*tag= az
FT      /note= "encodes K57"
FT      39589..40626
FT      /*tag= ba
FT      /note= "encodes A17"
FT      41341..41922
FT      /*tag= bd
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1,2e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GAGACCGAAGCGCGATGAGACGACCCAGCGC;CCGCGGTGGGTGCAGCCGACAGACGCGCGC 60
DB      28200 GAGACCGAAGCGCGATGAGACGACCCAGCGC;CCGCGGTGGGTGCAGCCGACAGACGCGCGC 28259

OY      61 CCGCGCGCGTTGCCGCGCTCGAGGTGCT;GAGAGCCCGCGC 101
DB      28260 CCGCGCGCGTTGCCGCGCTCGAGGTGCT;GAGAGCCCGCGC 28300

RESULT 3
ID      ABN25960 standard; cDNA; 384 BP.
XX
AC      ABN25960:
XX
DT      24-JUN-2002 (first entry)
XX
DE      Human ORFX polynucleotide sequence SEQ ID NO:20397.
XX
KW      Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW      hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW      degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW      cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW      hypertension; hypothyroidism; cholesterol ester storage disease;
KW      immune deficiency; immune disorder; infectious disease;
KW      autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW      myasthenia gravis; gene; ss.
XX
OS      Homo sapiens.
```

```
XX      WO200192523-A2.
PN
XX
XX      06-DEC-2001.
PD
XX
XX      29-MAY-2001; 2001MO-US10836.
PF
XX
XX      30-MAY-2000; 2000US-206132P.
PR
XX      29-AUG-2000; 2000US-228716P.
XX
PA      (CURA-) CURAGEN CORP.
PI
PI      Shinkets RA, Leach MD;
XX
XX      WPI; 2002-106308/14.
DR
XX      P-PSDB; ABP10208.
XX
XX      Novel human polypeptides and polynucleotides useful for diagnosing,
PT      preventing and treating cardiovascular disease, neurodegenerative,
PT      hyperproliferative disorders and autoimmune disorders
XX
XX      Disclosure: SEQ ID 20397; 1037pp; English.
XX
XX      The present invention describes substantially purified human proteins
XX      (referred to as open reading frame, ORFX, where x is 1-1191 (see Table 1
XX      in the specification). ABN15762 to ABN27252 encode the human ORFX
XX      proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX      treating or preventing a pathology associated with an ORFX-associated
XX      disorder in humans, and in the manufacture of a medicament for treating a
XX      syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX      sequences can be used in gene therapy. ORFX sequences can be used in the
XX      treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX      psoriasis, benign tumours, keloid, degenerative disorders, hemorrhage,
XX      osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic
XX      lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX      storage disease, various immune deficiencies and disorders, infectious
XX      disease, autoimmune disorders such as multiple sclerosis, rheumatoid
XX      arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX      disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX      useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX      bone degenerative disorders, or periodontal disease, and for gut
XX      protection or regeneration and treatment of lung or liver fibrosis,
XX      reperfusion injury in various tissues and conditions resulting from
XX      systemic cytokine damage.
XX      N.B. The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 384 BP; 75 A; 135 C; 98 G; 74 T; 2 other;

Query Match      32.3%; Score 32.6; DB 24; Length 384;
Best Local Similarity 58.9%; Pred. No. 14;
Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY      4 ACCGAGCGCGATGAGACCCAGCGCGCGGTGCAGCGACGACGCGCGCC 63
DB      262 AACGAGCGCGGCTCGGTGAGCCATGACACCGCGCGGTGCTCGCGGGGTGACGCG 203

OY      64 GCGCGGTGCCGCGCGCTCGAGGTGCTCGAGGACCG 98
DB      202 ATGCGCTTAGTGGGGTCTGCTGCGGGGTTTCG 168

RESULT 4
ID      AAL40781 standard; DNA; 88421 BP.
XX
AC      AAL40781:
XX
XX      03-OCT-2002 (first entry)
DT
DT      88421nt genomic DNA of ramoplanin producing Actinoplanes sp.
XX
DE
```

XX Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;
KW biosynthesis gene cluster; bioengineering; peptide synthetase module;
KW adenylation domain; hydroxyphenylglycine, HFG; antibiotic precursor;
KM chlorinate; lipdipsipeptide; gene; ds.
XX Actinoplanes sp.
XX Key location/Qualifiers
FH 2077..3078
FH /tag= a
FT /product= "Protein of ORF 1"
FT 3118..4032
FT /tag= b
FT /product= "Protein of ORF 2"
FT 4038..5048
FT /tag= c
FT /product= "Protein of ORF 3"
FT complement (6665..5814)
FT /tag= d
FT /product= "Protein of ORF 4"
FT complement (7703..6693)
FT /tag= e
FT /product= "Protein of ORF 5"
FT complement (9464..8130)
FT /tag= f
FT /product= "Protein of ORF 6"
FT 9691..10761
FT /tag= g
FT /product= "Protein of ORF 7"
FT complement (12751..10829)
FT /tag= h
FT /product= "Protein of ORF 8"
FT complement (13617..12802)
FT /tag= i
FT /product= "Protein of ORF 9"
FT complement (15203..13614)
FT /tag= j
FT /product= "Protein of ORF 10"
FT 15591..15863
FT /tag= k
FT /product= "Protein of ORF 11"
FT 15880..19035
FT /tag= l
FT /product= "Protein of ORF 12"
FT 19032..39713
FT /tag= m
FT /product= "Protein of ORF 13"
FT 39713..65800
FT /tag= n
FT /product= "Protein of ORF 14"
FT 65826..66530
FT /tag= o
FT /product= "Protein of ORF 15"
FT 66546..67370
FT /tag= p
FT /product= "Protein of ORF 16"
FT 67384..70059
FT /tag= q
FT /product= "Protein of ORF 17"
FT 70099..70662
FT /tag= r
FT /product= "Protein of ORF 18"
FT 70659..71906
FT /tag= s
FT /product= "Protein of ORF 19"
FT complement (73439..71964)
FT /tag= t
FT /product= "Protein of ORF 20"
FT complement (74216..73563)
FT /tag= u
FT /product= "Protein of ORF 21"
FT complement (75424..74213)
FT CDS

FT /tag= v
FT /product= "Protein of ORF 22"
FT 75535..76464
FT /tag= w
FT /product= "Protein of ORF 23"
FT complement (78110..76449)
FT /tag= x
FT /product= "Protein of ORF 24"
FT complement (79864..78107)
FT /tag= y
FT /product= "Protein of ORF 25"
FT complement (81624..79861)
FT /tag= z
FT /product= "Protein of ORF 26"
FT complement (81909..81682)
FT /tag= aa
FT /product= "Protein of ORF 27"
FT complement (82346..82062)
FT /tag= ab
FT /product= "Protein of ORF 28"
FT 82587..84446
FT /tag= ac
FT /product= "Protein of ORF 29"
FT 84481..85548
FT /tag= ad
FT /product= "Protein of ORF 30"
FT 85556..86845
FT /tag= ae
FT /product= "Protein of ORF 31"
FT 87372..86803
FT /tag= af
FT /product= "Protein of ORF 32"
FT 87494..88420
FT /tag= ag
FT /product= "Protein of ORF 33"
FT W020231155-A2.
FT 18-APR-2002.
FT
FT 15-OCT-2001; 2001WO-CA01462.
FT PF
FT 13-OCT-2000; 2000US-239924P.
FT PR
FT 12-APR-2001; 2001US-283296P.
FT PR
FT 24-JUL-2001; 2001US-0910813.
FT XX
FT PA
FT (ECP-) ECOPIA BIOSCIENCES INC.
FT Farnet CM, Zazopoulos E, Staffa A;
FT
FT WPI: 2002-435445/46.
FT DR
FT P-PSDB: AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,
FT DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,
FT DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,
FT DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,
FT DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.
FT XX
FT Novel isolated ramoplanin biosynthetic pathway polypeptide useful for
FT PT chemically modifying biological molecule that is a substrate for a
FT PT polypeptide encoded by a ramoplanin biosynthesis gene cluster -
FT XX
FT
FT Disclosure: Page 87-135; 212pp; English.
FT
FT
FT The invention relates to an isolated ramoplanin biosynthetic pathway
FT CC polypeptide selected from a polypeptide of open reading frames (ORF)
FT CC 1-32. The isolated polypeptides are useful for chemically modifying a
FT CC biological molecule that is a substrate for a polypeptide encoded by a
FT CC ramoplanin biosynthesis gene cluster, by contacting the biological
FT CC molecule with the isolated polypeptide, where the polypeptide chemically
FT CC modifies the biological molecule. The method comprises contacting the
FT CC biological molecule with at least two different polypeptides encoded by
FT CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the
FT CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated

CC gene cluster comprising the ORFs is useful as a substrate for
CC bioengineering of antibiotic structures. An isolated polypeptide or its
CC encoding nucleic acid sequence is useful for generating derivatives of
CC ramoplanin for improving production or for producing variants of other
CC antibiotics of the peptide class. The isolated polypeptides are useful
CC for synthesis of ramoplanin *in vivo* or *in vitro*, as an adenylation domain
CC in conjunction with other peptide synthetase modules and allowing the
CC incorporation of the into a peptide antibiotic precursor, for modifying the
CC fatty acid structure and/or enhancing fatty acid incorporation into the
CC (HPG)-containing peptide antibiotic, for production of an hydroxyphenylglycine
CC ramoplanin or its variants and derivatives, for enhancing uptake of
CC precursors for ramoplanin biosynthesis, for enhancing production of
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of
CC a peptide antibiotic precursor, and for designing specific lipopeptide
CC probes and primers for identifying and isolating putative lipopeptide
CC -producing microorganisms. This polynucleotide sequence represents the
CC 88421nt genomic DNA of a ramoplanin producing *Actinoplanes* sp.
CC microorganism of the invention.

XX Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 other;

Query Match 31.9%; Score 32.2; DB 24; Length 88421;
Best Local Similarity 59.1%; Pred. No. 13;
Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

XX
OY 6 CGAACGCCGATGAGCAGCCAGCCGCGCTGCGGTGCGACCGCAGACGCGCGCGCG 65
Db 78265 CGGCG 78206
OY 66 CGGCTGCG 98
Db 78205 CGGCTGCG 78173

RESULT 5
AAH05398
ID AAH05398 standard; cDNA; 829 BP.
AC AAH05398;
XX
XX 26-JUN-2001 (first entry)
DE Human cDNA clone (5'-primer) SEQ ID NO:2233.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
XX EP1074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -

PS Claim 1; SEQ ID 2233; 2537pp + CD ROM; English.

XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo or primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets are useful for antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
XX Sequence 829 BP; 145 A; 292 C; 239 G; 149 T; 4 other;

Query Match 31.7%; Score 32; DB 22; Length 829;
Best Local Similarity 60.2%; Pred. No. 19;
Matches 53; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

XX
OY 12 CCGATGACGACACCCAGCCCGCGGTGCGACGCCGACGAGCGCGCGCGCGCGTT 71
Db 94 CCGAGTGGCTTACCGAGCCCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 153
OY 72 GCCGCGCGCTGAGGTGCTGAGGAGCCGC 99
Db 154 GCGCGCTTGGAGGTGTGTGCGAGAGC 181

RESULT 6
AAH15833
ID AAH15833 standard; cDNA; 2210 BP.
AC AAH15833;
XX
XX 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:14325.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
XX EP1074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX

XX Key Location/Qualifiers
FH CDS 1..1128
FT /tag= a
FT /product= "GPCR 14266"
FT sig_peptide 1..102
FT /tag= b
FT mat_peptide 103..1125
FT /tag= c
FT /product= "Mature GPCR 14266"
XX
XX WO200212344-A2.
XX 14-FEB-2002.
XX
XX 08-AUG-2001; 2001WO-US24835.
XX
XX 09-AUG-2000; 2000US-0634392.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Welch NS;
XX
XX WPI; 2002-217181/27.
XX P-PSDB; AAEL14597.
XX
XX
XX Identifying compound that binds to and/or modulate activity of the
PT polypeptide for treating splenomegaly, emphysema and multiple
PT sclerosis, comprises using human G-protein coupled receptor 14266
PT polypeptide -
XX
XX Claim 1; Fig 2; 115pp; English.
XX
XX The invention relates to a method of identifying an agent that binds to
CC and/or modulates activity of human G-protein coupled receptor (GPCR)
CC 14266. The invention also provides a method for modulating the level
CC or activity of the GPCR in cells in vivo or in vitro, or in a subject
CC predisposed to having a haematopoietic or neutrophil deficiency
CC disorders. Detecting 14266 polypeptides or nucleic acids in specific
CC biological samples is useful in disease diagnosis. Modulating level or
CC activity of the GPCR polypeptide or polynucleotide is useful for treating
CC diseases related to 14266 receptor malfunction e.g., disorders involving
CC spleen (e.g. splenomegaly, neoplasms); lung (e.g. pulmonary hypertension,
CC emphysema); colon (e.g. diarrhoea and dysentery, idiopathic inflammatory
CC bowel disease); liver (e.g. hepatic injury, viral hepatitis); uterus and
CC endometrium (e.g. endometriosis, endometrial polyps); brain (e.g.
CC acute meningitis, multiple sclerosis, spinocerebellar degeneration);
CC T-cells (e.g. systemic lupus erythematosus, polyarteritis nodosa); skin
CC (e.g. vitiligo, acanthosis nigricans); heart (e.g. heart failure,
CC angina pectoris); blood vessels (e.g. atherosclerosis, Raynaud disease,
CC thalassemia syndromes, haemolytic anaemia) and disorders of the thymus
CC (e.g. thymic cysts, Hodgkin's disease). The present sequence is human
CC GPCR 14266 DNA.
XX
XX
SQ Sequence 1128 BP; 116 A; 437 C; 395 G; 180 T; 0 other;
XX
XX Query Match 30.9%; Score 31.2; DB 24; Length 1128;
XX Best Local Similarity 70.0%; Pred. No. 29;
XX Matches 42; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

0Y 32 GCGGCTGGAGCGAGCGAGCGCGCGCGCTGCGCGCTCGAGTGCTGG 91
DB 512 GCGGCTGGAGCGAGCGCGCGCGCGCGCGCTGCGCGCTCGCTGCTGCTGG 571

RESULT 11
ID AAS98049 standard; DNA; 1128 BP.
XX
XX AAS98049;
XX
XX 12-MAR-2002 (first entry)
XX

DE Human DNA for potential G protein-coupled receptor #7.
XX
XX Human: G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW tuberculosis; cognition disorder; memory disorder; anorexia;
KW hormonal release disorder; cardiovascular activity disorder;
KW pain perception disorder; obesity; diabetes; obesity;
XX diabetes; hyperlipidaemia; stroke; gene therapy.
XX
XX Homo sapiens.
XX
XX WO200185791-A1.
XX
XX 15-NOV-2001.
XX
XX 11-MAY-2001; 2001WO-US15332.
XX
XX 11-MAY-2000; 2000US-203217P.
XX 18-MAY-2000; 2000US-205945P.
XX
XX (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
XX Brown JP, Miller M, Burner G, Fabre-Suver C, Pritchard D;
XX WPI; 2002-066595/09.
XX
XX
XX Novel G protein-coupled receptor polypeptides including galanin
PT receptor polypeptides useful for identifying modulators that are useful
PT for treating Alzheimer's disease, psoriasis, melanoma, multiple
PT sclerosis, stroke -
XX
XX Disclosure; Page 116-117; 144pp; English.
XX
XX The invention relates to an isolated polypeptide encoded by a
CC nucleic acid molecule that is at least 80% identical to the G
CC protein-coupled (GPCR) polynucleotides included in the specification.
CC Also included are probes based on the GPCR sequences (including
CC antisense probes), a host cell comprising an expression vector comprising
CC the GPCR sequence, antibodies raised against the polypeptides,
CC and methods of identifying modulators of the polypeptides. The
CC polypeptides are useful for identifying modulator compounds which
CC function as modulators, activators, repressors, agonists or antagonists
CC of the novel GPCR polypeptides including the GAL4 polypeptide. The
CC antibodies and nucleic acid probes as described above can be used to
CC detect the presence of the polypeptides and nucleic acids and are used to
CC diagnose a variety of diseases or disorders in which GPCRs are involved
CC e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma,
CC atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy,
CC chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease,
CC depression, epilepsy, macular degeneration, lymphoma, melanoma,
CC multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease,
CC psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis,
CC tuberculosis and many other diseases listed in the specification. The
CC probes and antibodies are also useful for diagnosing cognition and memory
CC disorders, anorexia, hormonal release disorders, cardiovascular activity
CC disorders, pain perception disorders, obesity, diabetes, Alzheimer's
CC disease. Preferably, compounds that decrease or increase
CC the expression of galanin receptor (GAL4) can be used to treat obesity,
CC diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is
CC useful for treating the above mentioned disorders by gene therapy
CC techniques. The present sequence is a novel GPCR polynucleotide of the
XX invention.
XX
XX
SQ Sequence 1128 BP; 116 A; 437 C; 395 G; 180 T; 0 other;
XX
XX Query Match 30.9%; Score 31.2; DB 24; Length 1128;
XX Best Local Similarity 70.0%; Pred. No. 29;
XX Matches 42; Conservative 0; Mismatches 18; Indels 0; Gaps 0;


```

XX OS Homo sapiens.
XX PN WO200192523-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US10836.
XX PR 30-MAY-2000; 2000US-206132P.
XX PR 29-AUG-2000; 2000US-228716P.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach MD;
XX DR WPI; 2002-106308/14.
XX DR P-PSDB; ABP01922.
XX PT Novel human polypeptides and polynucleotides useful for diagnosing,
XX PT preventing and treating cardiovascular disease, neurodegenerative,
XX PT hyperproliferative disorders and autoimmune disorders
XX PS Disclosure; SEQ ID 3825; 1037pp; English.
XX CC The present invention describes substantially purified human proteins
XX CC (referred to as open reading frame, ORFX, where x is 1-11491 (see Table 1
XX CC in the specification). ABN15762 to ABN27252 encode the human ORFX
XX CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX CC treating or preventing a pathology associated with an ORFX-associated
XX CC disorder in humans, and in the manufacture of a medicament for treating a
XX CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX CC sequences can be used in gene therapy. ORFX sequences can be used in the
XX CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
XX CC osteoarthritis, neurodegenerative disorders, disorders related to organ
XX CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX CC storage disease, various immune deficiencies and disorders, infectious
XX CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX CC bone degenerative disorders, or periodontal disease, and for gut
XX CC protection or regeneration and treatment of lung or liver fibrosis,
XX CC reperfusion injury in various tissues and conditions resulting from
XX CC systemic cytokine damage.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 465 BP; 85 A; 165 C; 139 G; 75 T; 1 other:

Query Match 30.5%; Score 30.8; DB 24; Length 465;
Best Local Similarity 61.0%; Pred. No. 38;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 4 ACCGAAACCCGATGAGCAGCAACCCCGCGTGGTGACCCGACAGAGCGCGCC 63
DB 303 ACCGACCGCAGCAGCAAGCTGGCAGCACTCGTGCAGTGACCAACGCTTCGCTG 362
OY 64 GCGCGCTTCCGCGCGCTCGAGG 85
DB 363 TTCTCCCTGCGCGCGCGCGAGC 384

```

```

RESULT 15
ABQ90027/c
ID ABQ90027 standard; DNA; 1899 BP.
XX AC ABQ90027;
XX DT 01-OCT-2002 (first entry)

```

```

XX DE M. capsulatus gene #12 for DNA array.
XX KW Micro array; gene; ds; differential expression; gene expression.
XX OS Methylococcus capsulatus.
XX PN WO200255655-A2.
XX PD 18-JUL-2002.
XX PF 14-JAN-2002; 2002WO-NO00019.
XX PR 12-JAN-2001; 2001NO-0000235.
XX PR 12-JAN-2001; 2001NO-0000239.
XX PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX PA (TIGR-) TIGR.
XX PI Birkealand NK, Eidshammer I, Jonassen I, Jensen HB, Lien T;
XX PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
XX PI Salzberg SL;
XX DR WPI; 2002-557818/59.
XX PT Novel DNA array useful for determining differential expression of
XX PT Methylococcus capsulatus genes, comprises polynucleotides or
XX PT oligonucleotides representative for a selective number of Methylococcus
XX PT capsulatus genes
XX PS Claim 19; Page 52; 678pp; English.
XX CC The invention relates to a novel DNA array giving a representation of a
XX CC number of Methylococcus capsulatus genes. The method of the invention is
XX CC useful for determination of the differential expression of the genes of
XX CC M. capsulatus, and for studying gene expression on a genomic scale and in
XX CC gene expression assays of M. capsulatus genes. The sequences shown in
XX CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX CC invention.
XX SQ Sequence 1899 BP; 323 A; 580 C; 655 G; 341 T; 0 other:

Query Match 30.5%; Score 30.8; DB 24; Length 1899;
Best Local Similarity 58.9%; Pred. No. 36;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 4 ACCGAAACCCGATGAGCAGCAACCCCGCGTGGTGACCCGACAGAGCGCGCC 63
DB 294 ACCGCGGTGGCGGCTCGACAGCAACCCGCTGTAAGACAAATTGACGGCGAGCCC 235
OY 64 GCGCGCTTCCGCGCGCTCGAGG 93
DB 234 GTGAGCTTCGCTCCGCGAGCTATTGAG 205

```

Search completed: November 5, 2002, 13:38:47
Job time : 177.215 secs

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV1-FT0037-130
600-241-c066t3-2000-06-13&tl=4-1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 372.
Location/Qualifiers

FEATURES
source 1.372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0037"
/dev_stage="Adult"
/note="Organ: prostate tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 66 a 110 c 130 g 66 t

ORIGIN
Query Match 34.1%; Score 34.4; DB 12; Length 372;
Best Local Similarity 59.0%; Pred. No. 9.6;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 1 CGTCGCTCGGCTGCTGCTCTCTCGACCGCGCTGATCTGGAGGTTCAA 60
Db 133 CGTGGCAGACATCTCGGGATCCGCGACCTGCTGCTGATCACTGACGACCTGCGC 192
OY 61 TGAGCAGGCGCTGACTGCTGATGGCGTGGAGATCCGC 100
Db 193 GGACCTGGGCTGACTGCTCATGAGCGTGGAGTCCGC 232

RESULT 2
BE771151 37; bp mRNA linear EST 20-SEP-2000
LOCUS BE771151
DEFINITION RC5-FT0079-140600-031-H08 FT0079 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE771151
VERSION BE771151.1 GI:10224809
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 372)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brlones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, F. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC5-FT0079-140
600-031-H08&tl=2000-06-14&tl=1)
Seq primer: puc 18 forward
High quality sequence stop: 361.

FEATURES
source 1.372
Location/Qualifiers

1.372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="FT0079"
/dev_stage="Adult"
/note="Organ: prostate tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 66 a 115 c 127 g 64 t

ORIGIN
Query Match 34.1%; Score 34.4; DB 12; Length 372;
Best Local Similarity 59.0%; Pred. No. 9.6;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
OY 1 CGTCGCTCGGCTGCTGCTCTCTCGACCGCGCTGATCTGGAGGTTCAA 60
Db 185 CGTGGCAGACATCTCGGGATCCGCGACCTGCTGCTGATCACTGACGACCTGCGC 244
OY 61 TGAGCAGGCGCTGACTGCTGATGGCGTGGAGATCCGC 100
Db 245 GGACCTGGGCTGACTGCTCATGAGCGTGGAGTCCGC 284

RESULT 3
BG027149 378 bp mRNA linear EST 24-JAN-2001
LOCUS BG027149
DEFINITION BG0295826F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4390230 5',
mRNA sequence.
ACCESSION BG027149
VERSION BG027149.1 GI:12415488
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 378)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10079 row: d column: 07
High quality sequence stop: 378.
Location/Qualifiers

1.378
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4390230"
/clone_lib="NIH_MGC_86"
/tissue_type="osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bone; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: This is a NIH-MGC Library."

BASE COUNT 68 a 120 c 131 g 59 t

ORIGIN
Query Match 34.1%; Score 34.4; DB 12; Length 378;

Best Local Similarity 59.0%; Pred. No. 9.6;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 CGTGGCTCGGCTGCTGGGCTTCCTCGACCCGAGCGCGCTCATCTGGAGTGGGTTCAA 60
Db 48 CGTGGCACACATTCCTGGGACCTCCGACCTTGGCTGTCTCAACCTGGACACCTCAGTGGC 107

QY 61 TGACAGAGCGCTCGACTGCTGTGATGGCGGTGAGATCCGC 100
Db 108 GGACCTGGGCGCTGACTCGCTCATGAGCGGTGAGAGTGGC 147

RESULT 4
BE018358
LOCUS 393 bp mRNA linear EST 06-JUN-2000
DEFINITION bb79c01.y1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:3048576.5
Similar to gb:X13135 Murine mRNA fragment for fatty acid synthase
(MOUSE); mRNA sequence.
ACCESSION BE018358
VERSION BE018358.1 GI:8278381
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 393)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
imgc.lnl.gov/imap/html/fireources.shtml
Seq primer: -40bp from Gibco
High quality sequence stop: 383.
FEATURES
Location/Qualifiers
source
1. 393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3048576"
/clone_1id="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

BASE COUNT 71 a 119 c 137 g 65 t 1 others
ORIGIN

Query Match 34.1%; Score 34.4; DB 10; Length 393;
Best Local Similarity 59.0%; Pred. No. 9.7;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 CGTGGCTCGGCTGCTGGGCTTCCTCGACCCGAGCGCGCTCATCTGGAGTGGGTTCAA 60
Db 67 CGTGGCACACATTCCTGGGACCTCCGACCTTGGCTGTCTCAACCTGGACACCTCAGTGGC 126

QY 61 TGACAGAGCGCTCGACTGCTGTGATGGCGGTGAGATCCGC 100
Db 127 GGACCTGGGCGCTGACTCGCTCATGAGCGGTGAGAGTGGC 166

RESULT 5
BE018132
LOCUS 427 bp mRNA linear EST 21-SEP-2000
DEFINITION RC6-BN0276-160600-011-A12 BN0276 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE018132

VERSION BE018132.1 GI:10250366
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 427)
Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,
Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,
Goldman G.H., Carvalho A.F., Matsukuma A., Bata G.S., Simpson D.H.,
Brunstein A., deoliveira P.S., Bucher P., Jongeneel C.V., O'Hare
M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
Simpson A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LIRC Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=RC6-BN0276-160600-011-A12&l3=2000-06-16&l4=1>)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 418.
FEATURES
Location/Qualifiers
source
1. 427
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1id="BN0276"
/dev_stage="adult"
/note="Organ: breast-normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 83 a 129 c 144 g 71 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 12; Length 427;
Best Local Similarity 59.0%; Pred. No. 9.9;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 CGTGGCTCGGCTGCTGGGCTTCCTCGACCCGAGCGCGCTCATCTGGAGTGGGTTCAA 60
Db 197 CGTGGCACACATTCCTGGGACCTCCGACCTTGGCTGTCTCAACCTGGACACCTCAGTGGC 256

QY 61 TGACAGAGCGCTCGACTGCTGTGATGGCGGTGAGATCCGC 100
Db 257 GGACCTGGGCGCTGACTCGCTCATGAGCGGTGAGAGTGGC 296

RESULT 6
BF765829/c
LOCUS 430 bp mRNA linear EST 12-JAN-2001
DEFINITION IL2-CS0051-091100-223-D03 CS0051 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF765829
VERSION BF765829.1 GI:12113729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 430)

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mhc.mcl.nih.gov/
JOURNAL 1 (bases 1 to 510)
COMMENT Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM337 row: f column: 05
High quality sequence stop: 510.

FEATURES
source
1..510
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1="NIH-MGC-21"
/clone_11b="NIH-MGC-21"
/tissue_type="choriolealloma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 91 a 157 c 175 g 87 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 10; Length 510;
Best Local Similarity 59.0%; Pred. No. 10;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 1 CGTCGCTGGGCTGCTGCTCTGACCGAGCGCTGATCTCGGATGGGTCAA 60
DB 286 CGTGGACACATCTCGGATCGGACATCGGCTGCTGACACGCTCACTGCG 345
OY 61 TGACGAGGCGCTGACTGCTGATGCGGCGGATCCGC 100
DB 346 GACCTGGGCGCTGACTGCTGATGAGCGGGAGTGGCC 385

RESULT 12
LOCUS BE294484 545 bp mRNA linear EST 20-JUL-2000
DEFINITION 601173412P1 NIH-MGC_17 Homo sapiens CDNA clone IMAGE:3528986 5',
ACCESSION BE294484
VERSION BE294484.1 GI:9178021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mhc.mcl.nih.gov/
JOURNAL 1 (bases 1 to 545)
COMMENT Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM199 row: 9 column: 03
High quality sequence stop: 544.
Location/Qualifiers
1..545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1="IMAGE:3528986"
/clone_11b="NIH-MGC-17"
/tissue_type="pharyngosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 100 a 166 c 188 g 91 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 10; Length 545;
Best Local Similarity 59.0%; Pred. No. 10;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 1 CGTCGCTGGGCTGCTGCTCTGACCGAGCGCTGATCTCGGATGGGTCAA 60
DB 291 CGTGGACACATCTCGGATCGGACATCGGCTGCTGACACGCTCACTGCG 350
OY 61 TGACGAGGCGCTGACTGCTGATGCGGCGGATCCGC 100
DB 351 GACCTGGGCGCTGACTGCTGATGAGCGGTGAGTGGCC 390

RESULT 13
LOCUS BI084344 554 bp mRNA linear EST 20-JUN-2001
DEFINITION 602869861P1 NIH-MGC_102 Homo sapiens CDNA clone IMAGE:5014382 5',
ACCESSION BI084344
VERSION BI084344.1 GI:14502674
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mhc.mcl.nih.gov/
JOURNAL 1 (bases 1 to 554)
COMMENT Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1821 row: j column: 15
High quality sequence stop: 550.
Location/Qualifiers
1..554
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1="IMAGE:5014382"
/clone_11b="NIH-MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.

FEATURES
source

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submision
JOURNAL Submitted (16-DEC-1999) K'san Biosciences, Inc., 3832 Bay Center
place, Hayward, CA 94545, USA
FEATURES
source
1. 58733
/organism="Polyangium cellulosum"
/strain="Smp4"
/db_xref="taxon:56"
<1. 992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF61878.1"
/db_xref="GI:7453155"
/translation="VRCIVFAPADGCGPVLLYDNLKNAVYRRHDAIRPHITLAL
SDYREPRPVAVARCNLIKGRFRAIRYREGTEFERAIDLODLKQATENTSNAL
DSWYEDRAKTAQADIDRSVILKRPDIPFEPHEVEVGTPTARDLNTSPH
DTRRTLVYLAQSQRIADGNIVATHHNSMPGQITQEPHQLVDEKREARER
GIDRLKARKSSSAFLRIVAERDNGSAIRLQLDLVGALEAEALVDEPRTI
HIGAVQYIDRRSEHRI.PPVSIPIRGEHALVVTPTSHLTYDALKDPPT"
969. 1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLPTTEKDKLSLGLFGLACWEOIADKPFMLREVLAIEERE
RHKRLSERLKNRSVNAF.KPMTDFDSMPKIDREAVDDLYDSRYADLFEVVTTRYD
AQKPLSLSTNKAFAADWGQYFPHACVYTLVDRLVHRAVYLEAEASYRLKEAKELNAT
RTKQRTKKH"
1998. 6263
/gene="epoc"
/gene="epocA"
1998. 6263
/note="epocA"
/note="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEDSSAIAVIGMSGRPPGARIDETFRNLRDGTAVQSRSEQ
ELASGVDAVLADPNTYAGVLEVDVDFDAFGTISRRELMQPRITPMECAE
ALENNGDPTAEGSIGVAGNMSSTLSNHEPAMKRWGMQTLIGNDKDLAT
HVSIRNLNPGPSIVQATGSLVAVHLACMSLDREDCMALAGITVIRIPRAVYD
ABGIGTFSPDGHRAFDANKANGITMGCGVVLQVLEHGTOTLLGDILTEAALRVND
GARKIGTFAPSGVQAQALMEALALAGVBARIQVLEHGTOTLLGDILTEAALRVND
DSDASTRSCALGSKVTGIGHEASAGIAGLTKTVLAENHQLPSPLESPESPIDF
ASSPEYVNTSLDWTGTPPRRAGVSSGIGGTANHVYLEEAPPAKLPAAAPARSAL
FVVSAAALDPAARLRDLQAHQGLSLGVAFSPGSSQWVCMGRLLAMAPAREAL
RGCLDAAAGQTPPGAIVRGRCSPGVVPKVFYFPGSSQWVCMGRLLAMAPAREAL
SACDRAIOEAQMSGLAEALAGSSQIETRIDVQVYFALAVAPAAALMRSGVAPDV
VIGSHMGVEAAHVAGSLAEALAVVILICRSRLRLRISGOGMAVATELSLEAEAPLR
GYEDSVAVVNSPRSTVYSGPAAIAGVLSLANKGVYGRVAKVAVSQQVQDPLR
EDLLAAGGLRPRRAAVPMSRTVTCGAAVAGPLTANKGVYMMNIRQVAPAAVYQDPLR
CHGLFVENSHPHILITTSVEMRRAORAGVANSIRQDDEPAPALDELGLAMQGY
VFMGRLFPAGGRVPLPYTPMORERTYIEAPAKSACGDRGVRAGGHPLLGEMQLST

AVGLVSNAGLFPSSGGRVLPPTYPMQREYVIDTKADDAANGDRBRAGAGIDEYK
GAVRGDRRSARLDHPSPSGREVEAAGDPFLEIDEPEVLDPEVLRVYERAPG
LDEEYAVDAAGLSEFNDVQALGMPDDPGNPLRGDAGTACARIVAVGSGVGLV
VQDPYIALSNAGPAPHTVTSALVYDPRPALSATSAAMPVYLETAWALVYDSRSTAP
GRRVLIHAATGCGVLAANOMAHVGAAGTGBEKRAYLESIGVAVYDSRSTAP
ADVRATGCEGVYVNLNSCEGLIKSPFLHSRSPVIAKRDYCAQNOGSLRPEIR
NLSFSLVDRGMLEBRAPVRLPELITLACGYTPPTPTTLIARVADAFRMAQ
AHLGKVLITLDQPEVQIRIPTHAGGSGTGDRDLRLASAPAPAKALAEATYATQ
VSGVLRPEIKGAEALPRLGMDSLMAELNRIEASLKLSTLSTFSTSPNVLIT
QNLDLATALSLSERVAENLRAGVDSFVSSGADQDMWTAL"
6260. 10492
/gene="epoc"
/gene="epocB"
6260. 10492
/gene="epoc"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLENEHOGKILAADGERLOIOAPKNALPNILARISEH
KSTILMLRORLPASIVYAPAPRHHVPLPMDIQGSYMLGRTGATVPSGTHAREYD
CTDLVYARLSRAFRVVARHDLRAHTLPDMQVITEPYVDADIEITIDRGDRSTREA
RLVSLRDAMSHRIYDERPDLVAVVRLDEQOTRLVSLIDLINVLGSLSTIEPKMD
SPYDEPSTLPVLELSYRDYVALLESRRKSEAHOSMDYKRRVAVELPPPLMSASPEPT
BSTRIEIRPHREOMLPDSWSRLKORVGRGLTPGYVILAESFVIGRMSASPEPT
NITIRRLPVPHPRVNDITGDTSPSWLIDDTTRDKSPQRAKRIQEOLEAMNDHDS
GIENOREARVYGIORGALPVPVITSAINQVYGVTSIORCTGTPYVTSQTPQOLLDH
QLEHNDQVLAAMDVDCVYPPDILDMLEAVATLRITTEPSEORCSIPAOIE
NANSENLSLSEHRLGTFAPARQDPMQALVARSKTILEEELSRSLRSLAKRLK
EGARNTIYAVNMEGMEQVAVYALVDSGAVYPIIDDLDEAHLRLDGEKRLV
LQPMWLDGKVPFGIQRLLVSDAGVEGGDDLPMPQITQEDVLEITLGGTSLVPA
GVIMDRGAVNTLIDNERFEIGDGRVYLASSLSTSDVTSVTEFILAAGTIVGPA
SKUDLOAHAIIRPGLVYVMSVPALMMLVLEHSGRSDSLARSRLVSLGSDIAP
GLPGELOAIRPGVYVSLIGATVASTWSIGTVRVNDSPWASIPGRPLRQNTFVLD
EALDEPVPVWQKILYGGVLAGLGTWRDEKSRNSFLVHPETGERLYTGDDGLAVLP
GNIEMGRDNOIKLNGKVEJGEILETLKSPVNRADVTPVGDANAKMLLALVAP
EGTRRAEQDASLATERVADARAHAKAPAGSLDGRSCSLSVPEGGAALPKRRYSAG
TGLVPREAGLDVYARRSRVTLFAPITPVEFGRGLSCISVPEGRVAVQNDVPEDEA
FYLPEVGRIDALESRLGSLREFCILEGYMALIMEQASCNICVAVGQDFEQR
PVLDLHSDVYVHMGGRVDRPQOVCTIGODSPRRATTEGAPGCGDQPRADILRD
FLRTKLPXYMVTVVEJLDALPLTENGKDRALKERDTSPPRSGHTAPADALDEI
LVAAYREVYGLDEVQAOSEFVLDGATSIHIVMSRLQKRLDREITAITELFOYPMIGS
LVSGLRDSKDEQRPNMQDRVEARRKGRRS"
10639. 16137
/gene="epoc"
/gene="epoc"
10639. 16137
/gene="epoc"
/note="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEDSSAIAVIGMSGRPPGARIDETFRNLRDGTAVQSRSEQ
ELASGVDAVLADPNTYAGVLEVDVDFDAFGTISRRELMQPRITPMECAE
ALENNGDPTAEGSIGVAGNMSSTLSNHEPAMKRWGMQTLIGNDKDLAT
HVSIRNLNPGPSIVQATGSLVAVHLACMSLDREDCMALAGITVIRIPRAVYD
ABGIGTFSPDGHRAFDANKANGITMGCGVVLQVLEHGTOTLLGDILTEAALRVND
GARKIGTFAPSGVQAQALMEALALAGVBARIQVLEHGTOTLLGDILTEAALRVND
DSDASTRSCALGSKVTGIGHEASAGIAGLTKTVLAENHQLPSPLESPESPIDF
ASSPEYVNTSLDWTGTPPRRAGVSSGIGGTANHVYLEEAPPAKLPAAAPARSAL
FVVSAAALDPAARLRDLQAHQGLSLGVAFSPGSSQWVCMGRLLAMAPAREAL
RGCLDAAAGQTPPGAIVRGRCSPGVVPKVFYFPGSSQWVCMGRLLAMAPAREAL
SACDRAIOEAQMSGLAEALAGSSQIETRIDVQVYFALAVAPAAALMRSGVAPDV
VIGSHMGVEAAHVAGSLAEALAVVILICRSRLRLRISGOGMAVATELSLEAEAPLR
GYEDSVAVVNSPRSTVYSGPAAIAGVLSLANKGVYGRVAKVAVSQQVQDPLR
EDLLAAGGLRPRRAAVPMSRTVTCGAAVAGPLTANKGVYMMNIRQVAPAAVYQDPLR
CHGLFVENSHPHILITTSVEMRRAORAGVANSIRQDDEPAPALDELGLAMQGY
VFMGRLFPAGGRVPLPYTPMORERTYIEAPAKSACGDRGVRAGGHPLLGEMQLST

Query Match	Best Local Similarity	Score 101;	DB 1;	Length 58733;
Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	GC	GC	GC	GC
Db 25850	GC	GC	GC	GC
QY 61	GC	GC	GC	GC
Db 25910	GC	GC	GC	GC
RESULT 2	AF210843	68750 bp	DNA	Linear
LOCUS	AF210843	68750 bp	DNA	Linear
DEFINITION	Soorangium cellulosum strain so ce90 epothilone biosynthesis gene			
ACCESSION	AF210843			
VERSION	AF210843			
KEYWORDS	AF210843.1			
SOURCE	polyangium cellulosum			
ORGANISM	Bacteria; Proteobacteria; delta subdivision; Myxobacteria;			
REFERENCE	Myxococcales; Soorangineae; Polyangiaceae; Polyangium.			
AUTHORS	1 (bases 1 to 68750)			
	Molnar, I., Schupp, T., Ono, N., Zirkle, R., Milamow, M.,			
	Nwak, Thompson, B., Engel, N., Toupet, C., Stratham, A., Cyr, D. D.,			
	Gorlach, J., Mayo, J. M., Hu, A., Goff, S., Schmid, J. and Ligon, J. M.			
TITLE	The biosynthetic gene cluster for the microtubule-stabilizing			
	agents epothilone A and B from Soorangium cellulosum So ce90			

JOURNAL	Chem. Biol.	7 (2), 97-109 (2000)
MEDLINE	20130945	
PUBMED	10662695	
REFERENCE	2 (bases 1 to 68750)	
AUTHORS	Molnar, I.	
TITLE	Direct Submission	
JOURNAL	Submitted (03-DEC-1999) Natural Product Genetics, Novartis Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box 12257, Research Triangle Park, NC 27709, USA	
FEATURES	location/Qualifiers	
source	1..68750	
	/organism="Polyangium cellulosum"	
	/strain="So ce90"	
	/db_xref="taxon:56"	
	<1..1826	
CDS	/note="ORF1"	
	/codon_start=3	
	/transl_table=11	
	/product="unknown"	
	/protein_id="AAE26904.1"	
	/db_xref="GI:6724238"	
	/translation="ASIDLAFATARSARVLDGCHATERHVLAEGADIEDLRALREH LRIQEGSPHMCIGDLYVEELADOLPLASISPHNARSIRHPMTSDALVOCPLV RMIAAGAPGLPEREERERARACORARIMTAAPCPARLPREDNADLGGP MSPEVAERRLRASTVPELACALALWILCTGAGPSGIPATMDPENLILGELPT ALAAASGPTSEALRGAARLFAFWSEVSSKSQLGNI PALMERLETVIRANGNADN LSRERAEALAEARLRRAOPAPFAAGLAAGVSSSGRLSGLYTDGALYSGDND IVMPGRISIPVVLVAGTDEPPELAPLSOMLVAHANACTISKVLEGSPLIYMANR OARMSI.VHARGFWAMVNOAMPDPDEGAPVYVORSTIMEEHPPTCLHEPGASARS LACDEHLWYCELSAGRLLELMRHHRRGAPSRATYIGHPRIATWYPSLTIANHYL WADDRRAILIGVDKRTGVEPTVLAERHPAHVSEDRDLPALGDPDSKDMVHEIR SGASTVADYQROLDMPRDMVLNRRGLFEFTNRIILTLAS"	
CDS	/note="ORF2: similar to hypothetical proteins from Mycobacterium tuberculosis and Streptomyces coelicolor and to D2-peptidases"	
	/codon_start=1	
	/transl_table=11	
	/product="unknown"	
	/protein_id="AAE26905.1"	
	/db_xref="GI:6724239"	
	/translation="MGALISVAARGCALGGAEBEGQPGODAGATAPAREMAEVA AGQDGDSPVDRMLPELANRKLVARIDGP IDEVPAERPIVPHDLMTFTGEGISFDA SSPJORADIGELVNAQPVPMTPHGPDEMTIRLGTPLAMQKONYNNGSLVOGVL VGRRAIDGDFAFVERIERIAPLGMRDYDFHPADKLAFACGCFETDEOTEKTRMDH GAEVAVSPAPFSGAGLVSTVDYDILFARMIMGVHGGRLLSAASYEMTADHL TPACKAASFFPCFETGHMGYCAVYTADAYSEVTERIGMDGPGFTSMINDPGRIL IGIVMTDSAGLEFGALERFWSYVATESA"	
CDS	3415..5556	
	/note="ORF3: similar to Na/H and drug/H antiporters"	
	/codon_start=1	
	/transl_table=11	
	/product="unknown"	
	/protein_id="AAE26906.1"	
	/db_xref="GI:6724240"	
	/translation="MGCTIEROVLVLTVALILVTPARASGELARRLQPEVIGLEFG GVVLPESVGAALAPGFHRLPEBPVAVGVSGISGAILLMLAGIEVNOGILIRKA RPPALSALGAIAPRIAGCAFSALYIDRPLPSGLFTGIVTSVAATYATILERSBM RRTSAOVTTLAAGVSEVAANVLAMTSSGASBALAVANSALLASGFLLPVIVGR LTHLARMVADATRVSGQVSLVLTFLAALITQRGLHPLGAELVGLVLSNPTR NRPILDGVQTLVAGLFAPVFFVLGAMVDVDSQRTPARMGTVALLTATANAAYVPA LGAELDILGRSEALVAVGLNMKGDTLVAIVGELGLSNELATMYAVVALVTVTA SPALLWLEKRAPTOESARLEREEARARAYIPGVERILVPIVAHLPEDFADIES IVASKRKTETVITTELSTVFOQAPGSPRAGCEASRGILRGARLQVCIIMORIELGS IQALIRSRDHLVIGARSPEARARGSGFERLDATVORASENVLYVYGDPPAERAS ARRLVPIIGLEYSEFAADLAHAHVALAMDEVLVLSAQTPGAPVAVWRDPSRIVAV ARSVAEAVRGGRRLGVRSRVHVGHPDEDTRETLARAPYDLVLVGYCDHBPGLRT YLSTGSEVYVRSVLPALLVAGGTREQR"	
CDS	complement(5612..5992)	
	/note="ORF4"	
	/codon_start=1	
	/transl_table=11	

/product="unknown"
/protein_id="AAE26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGRTRCAIEGYIPGSGNGPEPOMTSHETACLLNADSDRA
OYATVYFSDRDPAGPYRVTPVAPRRTRHFRNDLPEPIFPRDTDVASVIESDAPIV
OHTRLDSROAEVALLTIVATDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAE26908.1"
/db_xref="GI:6724242"
/translation="MHVDGRRFRGRIHTGLASMAAGCGPSEKTVGCTRLAP
GDARATADVDDAKITRLADVHLSPERLEAGSERVVMORSPESPMMRVGLD
TKADSRGLAETVYANFELLITAEKSSPQSPSSAAVIGTISVG"
7610..11875
/gene="epoa"
7610..11875
/note="epoa"
/gene="EPOS A"
/codon_start=1
/transl_table=1
/product="polyketide synthase"
/protein_id="AAE26919.1"
/db_xref="GI:6724253"
/translation="MALRIERAEDPIAIVGASGRLPGVIDLSQFMTLLGSRDVT
GVPAERMDAAAFDPIPDAPGKTPTVTRASFSDVACFPDASFFGISPREALBMDPAH
LLEVCMEALENAAIAFSAVGTETGVFIGIPSEYALPQATASAEIDAHGIGTM
PSVGARISYALGRGCVAVDTAVSSLVAVHACQSLRSGECSTALAGGSLMSP
STLWMLSKTRALARDGECFAEADGFGRGCAVYVYKRLSGARADGDRILAVIRG
SAINDGASSGLTPNGSOSQEIWKRALDAGCAATGVTGVEAHGTGTLGDTITOA
LNAVIGIGRDVATPELLIISVKTNLIGHPSAGITGLIYVLSIOHGOITPAHHAALN
PRIWGDLRIYTRARTWPDNTPRAGVSSFGMSCTAHAYLEAPACNCTPAP
REALLVLSARTASALDQAARLDHLETGSPSCGDAFSLATRSANENHIAAAT
SREGIPLAALDAAGCTRGAVRSADDSKRLAFLPTGCGNOTLGGGLIDVNSAF
REAPLCVRLFNQDLDRUYREYMAEPASVDNALDQTFNTPALFETIALALMS
NGVEPELVAGHSIGELVILACVAGVPSLEDAVTLVANGTLMQALPRGGMVSTENEA
DVAAPVAPHAASVSTANAPDOVVIAGAGVTHAIAAMARAGATKLSHAFHS
PLMAHLEAGREAVSRRPSTIVLSAGKACTDEVSSEGVVRIHAREVRRADG
VALHAGAGTVEVGPSTLLGLVAPACPNRACITDEVSSGVVRIHAREVRRADG
AVGLVSMAGLTPSGRIPLPTVPMQRRKVIIDTKADDAAGDRRAPAGHDEVEEG
GAVRGGDRSARLDHPLESGRREKVEAAGDRFRLDEPEVLHDLVIRVERAPG
LGEVIVADAGLSNDVPLAGMWPDPGRPNPLLDPEVLDHLVIRVERAPG
VGPVIALATGAGVATHVITSAVLVLRDPLASAIEMAAFPVYLTAWYLDRIARLP
GERVIAHATGAGVLAAYMAOHVGAEVATAGTEKRAVYLSGVRYSDRSDFV
ADVRAMTGGEGVDVUNLSLSELIDKSFNLLRSRFTVGLGRDCYADQGLRFLR
NLSEFLVLDLGMMLERPAIVRALLEELGLIAGVFTPPITLPIARVADAFRSMAQ
AOHLKRLVLTLDPEVOIIPTHAGAGPSTGDRDLDRLASAAPARAALALPFLRTQ
VSQVLRTEIKVGAELFRLGMDSLMAVELNRIEASIKLITSTTFLSTSPNIALA
ONLIDALATLLELREVAIINLRAGVONDEVSSGADQDMETIAL"
7640..8920
/gene="epoa"
/note="Region: betc-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
/gene="epop"
11872..16104
/gene="epop"
/note="EPOS P: contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=1

/product="nonribosomal peptide synthetase"
/protein_id="AAE26925.1"
/db_xref="GI:6724259"
/translation="MTINOLNLEHOGIKLAADERLOIQAOKNALPNLLARISEH
KSTITLRLRORPAESIVAPAEHAPPELDIDIOSVYLGRTGATVSGVHAREYD
CTDLVPLRSARFKEVVARHMLRAHNTLPDMQVLEPVVDIETLIDRLGIDRSREA
RLVSLRDMASHRIYDTERPPLVHVAVASERORPVLISLINVDLGSLSIIRKMD
SPYEDPSTLYLETSDRYVALAESRKSSEHOSMXYMKRRALALPEPPITPKAD
BSTKEIRPRHREOMLPDSNGRLKRRVCEGCTPTQVILAFSEVIGRMSASPERT
NITIFNRLPYHRRVNDITGPTSLIIDIITDTPKSPQAKRTOEOLRPMMDCHVS
GTEVORPAAVIGTORGLFPVULTSALNOQVYTSQRIGTPVYTSQTPOLLDDH
QTERHGDVYLAAMDVDEFPPLDDMLDMLVYFLRLUTEPMEOYRGSLPRAOLE
AKASANNALISEHTLNGLRARVEQLPQOLAVSARKTLTELSRSTRKRLR
EOGARNTLVAVMKEMEOYAVLAVLESGAAVPIQDAPDAERTHVLIDHGEVGLV
LQPMWLDKLSMPGILQRLVSEAGVEDQPPMP1QTPSDALVAVIYTSGLPK
GVMIHRAVAVNTIIDINERFETGPDRLVALSSLSFSDVYDVGILAAAGTIVPDA
SKLRDPHMAVELIEREKVTVMNSVPALMRMLVEHREGRPSDLARLSRLSGMIVP
GLPGLQAIRGVVISLGATBASIMSIGVPRNVDSLMSIPYGRPLRNQTEHVL
BALERPVPWPGQVIGVGALAGYWRDEKTRKSLVHPETGERLYKTDGLRYLP
GNIEPMGEREDNOIKLRGRVLELGEIETLKSHPNRDAVIVPVGNDANKLLAVYV
EGTRRRAAEODASLKTERRIDARAHAAEADGSDGERVOFKLARHOLRDLGKPVVDL
TGODPREAGLDVYARRSVTRTELEAPIDPEFVEGRSLCSEVPEPGATLPKFRYPSAG
SIYPVQTYAVVKSRIEGVDEGFYVYHPEHRLTLKLSHGIERGCAHVQNDVDEAA
FULLLRHSDVYVHGMIGRVPDPOVCTLODSSPRATRGAPGSGREOHFADLRD
FLRTKLPEYVPTVEVDALPLTNGVDRKALREKDOTSPRHSRGTAPDRLAEI
LVAVREVILGEVGLQOSFVOLDGATSIHYMRSLQKRLDREIATITELQVPMIGS
LASGLRDSRDLDQRPNNQDRVEYRKRGRRS"
16251..21749
/gene="epob"
16251..21749
/note="EPOS B"
/codon_start=1
/transl_table=1
/product="polyketide synthase"
/protein_id="AAE26920.1"
/db_xref="GI:6724254"

Query Match 100.0%; Score 101; DB 1; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3; le-11;
Matches 101; Conservative 0; Mismatches 0; Gaps 0;
OY 1 GCGCGGCTTTCGCCCGACGGCGGTGCAAGACGTTCTCGCGCGACGCGGCTACGC 60
|||||
Db 31462 GCGCGGCTTTCGCCCGACGGCGGTGCAAGACGTTCTCGCGCGACGCGGCTACGC 31521
|||||
OY 61 GCGGGGCGAGGGGTGCGCGCTGCTCAAGCGCGTGC 101
|||||
Db 31522 GCGGGGCGAGGGGTGCGCGCTGCTCAAGCGCGTGC 31562
|||||

RESULT 3
AR193029
LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn, and
Goriach,J.
TITLE Genes for the biosynthesis of epoethaliones
JOURNAL Patent: us 6346404-A 1 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..68750
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORGANISM "unknown"

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 31462 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 31521
|||||

Qy 1 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 60
|||||

Db 31462 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 31521
|||||

Qy 61 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 101
|||||

Db 31522 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 31562
|||||

RESULT 4
ARI99551
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 60
|||||

Db 31462 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 31521
|||||

Qy 61 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 101
|||||

Db 31522 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 31562
|||||

RESULT 5
ARI99559
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 60
|||||

Db 31462 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 31521
|||||

Qy 61 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 101
|||||

Db 31522 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 31562
|||||

Db 31462 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 31521
|||||

Qy 61 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 101
|||||

Db 31522 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 31562
|||||

RESULT 6
ARI99567
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 60
|||||

Db 31462 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 31521
|||||

Qy 61 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 101
|||||

Db 31522 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 31562
|||||

RESULT 7
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 60
|||||

Db 31462 GCGGCGCTTTCGCCGCGAGCGGCGGTGCAAGACGTTTCGCGCCGACGGGAGCGGCTACGC 31521
|||||

Qy 61 GCGGCGCGAGGGGTGCGCGCGTGTGCTCAAGCGGCTGC 101
|||||

Db 31522 GCGGGCGAGGGGTGCGCGCTGTGGTGCCTCAAGCGGCTGC 31562

RESULT 8
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Molnar,I., Zirkle,R., Cyr,D.,Dawn, and Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCGCGGCTTTGCGCGGAGGCGGTGCAGACGTTCTCGCGGAGCGGCGCTACGC 60
|||||
Db 31462 GCGCGGCTTTGCGCGGAGGCGGTGCAGACGTTCTCGCGGAGCGGCGCTACGC 31521
|||||

Oy 61 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 101
|||||
Db 31522 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 31562
|||||

RESULT 9
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C., and Tang,L.
TITLE Recombinant methods and materials for producing epoethlones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source Location/Qualifiers
1..71989 /organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 3.1e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCGCGGCTTTGCGCGGAGGCGGTGCAGACGTTCTCGCGGAGCGGCGCTACGC 60
|||||
Db 25850 GCGCGGCTTTGCGCGGAGGCGGTGCAGACGTTCTCGCGGAGCGGCGCTACGC 25909
|||||

Oy 61 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 101
|||||
Db 25910 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 25950
|||||

RESULT 10
LOCUS AX024383 10910 bp DNA linear BCT 15-SEP-2000

DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024383
VERSION AX024383.1 GI:10184587
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum.
REFERENCE 1 (bases 1 to 10910)
AUTHORS Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
Bayer,S., and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source Location/Qualifiers
1..10910 /organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 1852 a 3124 c 3613 g 2321 t

Query Match 79.4%; Score 80.2; DB 1; Length 10910;
Best Local Similarity 87.1%; Pred. No. 5.8e-07;
Matches 88; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 1 GCGCGGCTTTGCGCGGAGGCGGTGCAGACGTTCTCGCGGAGCGGCGCTACGC 60
|||||
Db 2897 GCGTTGCTTTGCGCGAGATGGGCGGTGCAGACGTTCTCGCGGAGCGGCTTGC 2956
|||||

Oy 61 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 101
|||||
Db 2957 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 2997
|||||

RESULT 11
LOCUS AX024276 10910 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024276
VERSION AX024276.1 GI:10184550
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum.
REFERENCE 1 (bases 1 to 10910)
AUTHORS Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
Bayer,S., and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source Location/Qualifiers
1..10910 /organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 1852 a 3124 c 3613 g 2321 t

Query Match 79.4%; Score 80.2; DB 6; Length 10910;
Best Local Similarity 87.1%; Pred. No. 5.8e-07;
Matches 88; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 1 GCGCGGCTTTGCGCGGAGGCGGTGCAGACGTTCTCGCGGAGCGGCGCTACGC 60
|||||
Db 2897 GCGTTGCTTTGCGCGAGATGGGCGGTGCAGACGTTCTCGCGGAGCGGCTTGC 2956
|||||

Oy 61 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 101
|||||
Db 2957 GCGGGCGAGGGGTGCGCGCTGTGTGCTCAAGCGGCTGC 2997
|||||

RESULT 12
LOCUS AR166425 33529 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 3 from patent US 6280999.
ACCESSION AR166425
VERSION AR166425.1 GI:16241741

KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 33529)
TITLE	Gustafsson, C., Betlach, M.C., Ashley, G., Julien, B. and Ziermann, R.
JOURNAL	Soraniom polyketide synthases and encoding DNA therefor
FEATURES	Patent: US 6280999-A 3 28-AUG-2001;
SOURCE	Location/Qualifiers
	1..33529
BASE COUNT	/organism="unknown"
ORIGIN	4489 a 9518 c 14470 g 5046 t 6 others
Query Match	64.8%; Score 65.4; DB 6; Length 33529;
Best Local Similarity	78.8%; Pred. No.0.00037;
Matches	78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
OY	1 GCGCGCGCTTTCGCCCGACGGGGGTGCAAGACGTTCTCGGCCCGACGGGACGGCTAGGC 60
Db	10351 GCGTCCGGGTCTTCGCCCGACGGAGCGGTGCAAGCGTTCTCGGACGGCGCCGACGGCTACGG 10410
OY	61 GCGGGGCGAGGGGCGTCCGCGTGTGCTCTCAAGCGGCT 99
Db	10411 CCGCGCGCAGGGGCGTCCGCGGTGTGCTCTCAAGCGGCT 10449
RESULT 13	
AF188287	42603 bp DNA linear BCT 12-JAN-2000
LOCUS	AF188287
DEFINITION	Stigmatella aurantiaca myxothiazol gene cluster, complete sequence.
ACCESSION	AF188287
VERSION	AF188287.1 GI:6635393
KEYWORDS	Stigmatella aurantiaca.
SOURCE	Stigmatella aurantiaca.
ORGANISM	Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
	Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE	1 (bases 1 to 42603)
AUTHORS	Slakowski, B., Schallier, H.U., Ehret, H., Kunze, B., Weiniß, S.,
	Nordstiek, G., Brandt, P., Bloeker, H., Hofle, G., Beyer, S. and
	Müller, R.
TITLE	New lessons for combinatorial biosynthesis from myxobacteria. The
	myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca
JOURNAL	DM4/3-1
MEDLINE	J. Biol. Chem. 274 (52), 37391-37399 (1999)
PUBMED	20069734
REFERENCE	10601310
AUTHORS	2 (bases 1 to 42603)
	Slakowski, B., Schallier, H.U., Ehret, H., Kunze, B., Weiniß, S.,
	Nordstiek, G., Brandt, P., Bloeker, H., Hofle, G., Beyer, S. and
	Mueller, R.
TITLE	Direct Submission
JOURNAL	Submitted (20-SEP-1999) NBI/MX, GBF, Mascheroder Weg 1,
FEATURES	Braunschweig 38124, Germany
SOURCE	Location/Qualifiers
	1..42603
	/organism="Stigmatella aurantiaca"
	/strain="DM4/3-1"
	/db_xref="taxon:41"
	1..834
gene	/gene="mtaA"
	1..834
CDS	/gene="mtaA"
	/note="puta transferase"
	/codon_start=1
	/translation="MPPSPALPLIKPPPEVHWIYPERITPEGLLESYRALDGG
	ERDKQAFYFERHRLQYVSHALVRLTSLRAVAPAWSSANQYGPETRGEEKFW
	LRFNLSFDNALCAVARVDVGDVADVEDTERGFTVIAUSFAPAPAVASIRALPVSG

ORERFDYDYLKEATYIKAKGMGSLSLDOPAFVSGLSIRISFDRIDVEDSOMOD
RPRPSORHAAALAVRPSDAPLTVRRQRTYRPLDDDAPELTSERIQRLRMRPGVG
G
1388. .13399
/gene="mtab"
/note="PKS"
1388. 13399
/gene="mtab"
/note="PKS; Mtab loading domain and module 1: polyketide
synthase domains: ACP, KS, AT, Ar, DH, ER, KR, ACP; Mtab
module 2: polyketide synthase domains: KS, AT, DH, KR,
ACP"
/codon_start=1
/transl_table=11
/product="Mtab"
/protein_id="AAF19810.1"
/db_xref="GI:6635395"
/translation="MSKEHGVCVGVSGTGLRPMYLVOOVAQSVGLEAHNRITPEFS
LGVGSPFATRIILAEKARLIERLSPITLVNHPHITIDKFAHRLSNRMDASLYLDPAGTP
ADGAEPIAIVIGLACEFPAAPDAKAWRYLOBERCDVTEVPNSNRMDASLYLDPAGTP
GRTIVYGFGLDSVDFDFVEFQGISREGQIDPQORLMELSMWELADGIAPIRYLR
SGTIVGTGEMWMDYADLOTILHAAPIHLITCTAGGILANRISYALGEGPSLADIV
ACSSSLVAIHEACRSLSRSGSSSLTAGVOLINSPRPMVAFGALAPRGCRTEPELA
RASGVYRGCGGVYLYLKRSLALDDAPYICLRGSANVNDGASNGLAPRACRTEPELA
LNDIAVAMDDQWHTVHELHGITGLDPTTEKAGALVIGARREARPLHVSQKTN
IGHLAAADIASTYKVALSIRHQRDASLHEFQENPRTIPDDILRATVOTLAFMAPAK
DPAVGIISAFEGCGTCHVHEHREHSGSRATPAVNPSPHRTALPLFAGASGA
AMGSGEHLGAEPVFAFEECDAPRGLGMSGLVALLARVSRGCEQERDPLVLA
ARQISLAALMRBSGTEIPSNYDQSVEMBAAOVAGLITIEVVALRYHYOVQKGVN
GYGMAYVIGSGAEVARINAPYGRGISIASENGSRSTGSGEAAALDFTASMAAGF
CYMAVEIDFAHLPVEVALPLPLFQSCVLPKSPGRPMVSTVSGAFVDDTGLAEHWA
SNREDEVPATIMRTLIEGGYDVLVDNPHPIHEHSLVENVRGVTSAEETIASMRDEE
PVEGHLHRLTOLSOAGGVPMRVRVVPVGAGSSAEVRELFYASATBOALREOARD
IAEORDDAASLTIDCTYAGARSHHEHRLVAETRALEEGELAFARGELTPNPS
QGMVRGEARPLFGVSGGSPGLSSQVFEESPRKAFESERLFGNOTASLIE
ALHABESQHPRTTEPOLYLPARFOVALALMSKGRIDPAVYGHSTGELSAANVAG
LSLEELSLVYSRTRMGRAGGAGVATVBSARTEPLRGLCGTDLATAVNPENVA
VYSGALAAALASLDELITOGGLRVLRYLVNVAFFASAOAPYLGOLDVRLRKRPPRL
PIYSYHGHLGASAPDGTIDMRANVEREYLPAAVAMRADGCAFEVETSPHVLQDLP
LSRIDPSSAGSALFSRDRADPAALLNGLTLYTLGHEBSMKVPRDLGCTVEL
PEYPMORQCMFMDAPOLPAEPHRLPGLRVYDAPEGQVEALFEPSRQDLODHRIL
GVYVAGASHLGMVLVAANQMGAEACALODYLPFLPYLGTIEDRLVOYSTINDAF
RILSNAGQGGEGESHWITAHATRGLRGKEVETDAGEVLSVEYARPLASQVLE
GIARQGLGROFRMLAEALMOGRREAFGRRLABETDGDADYFHPRLIDSCQFTE
VAGFSIEGLAEVPLGVGFALGAEARPRFCLVYRSGGKANESASVDRLIDESGG
LVAAEVSGLTVIRAPREPELRYVRYAPLPREFPMOPQAOVEALFEPSRQDLODHRIL
GLAARLABDSSRGSVEYVLAEENMOAGTREALDASLACAGVYDLGDEGSS
DAVQGNARVQDLVTRPAADALRLVLYTRKQTVSGDSSIAEQARLKLGVATLE
HELRQRIIDLPSPRAADALAEALAPGRDVAFRGRRLVAETVKTLPPVAVR
ALPLEDAVLYLVGGLGGIGLEIARMMWGASGHLVLMGSSASAHQOELALRDAG
ALISVALGVNVEELAEALVIGLDIRAMPRLRGMAHLYGDDCLVQOSERIALRYLA
PIQCGMMNHLREYQGRDLPFPGKSSVSSVSGGQAGVAAANAPLIDALSLHRAQGL
TALSTWGMQWADYGMARLPPEAGKRWSSFGGLDLPASAGLALRILSLSEANOVYVA
GASVEAKQDFAFEDPALSRSKRKESAPVSGVALGASRELRARLUSLQVATIV
AGLPPSAPIDPRPPLQGLDLSLMALQFRLSIASLSGRSLPALLREDPHLELLEFL
ADLPPLPASASPAKSRRPYLPARDAEGEAPLAVVMACTLSVSPEPRFWELRLANG
VQARLEPDEKMSLIDWYDPDPTPEKXSRHGCFGLDEVDPEPAFFGSIAPARASGD
POORITLLEGMELWESAGLVAIDOLKSPVTEGLCISDVALLELNSPDTGFINASGD
SGGYSVAGARLAIYALGEGPLTADYPTACSSIVAAHLACOSLRNECDALDYGSL
LISPRASVFSKLALSDPGACKADPSSAGVYRGSGAGVLYLKRSEALADDPILSL
VNRGVAVNDGSGNSLTATSAQOENPISALTKQGVSPDVSYLEHAGETSLGDT
EVOALAAVYAGKRPRAPIGSGVNTKVLBECAATYSIKALILRHLREPLSLIF
KPPSAPIMDELPREAPYIESHRPVRVEGRKRLIGVSFGSGNANVILEADACRSL
SKAPSIPREBELLYISARPAOLOEAENAPPAIPISYTVAGKAENDSYMANRLRO
VLAPEVLTILAEEDYTLPELQGNAPPAIPISYTVAGKAENDSYMANRLRO
RSLVGGPLTAQVTPAMNRLPPEBGSVLSLPSTYMWOKERYIASPTARQDAPSGCV
VRLADIVFPRPLVLEKGVENKMPRLREHLHGALVAGAGVAAALASLRETRGTA
VRLADIVFPRPLVLEKGVENKMPRLREHLHGALVAGAGVAAALASLRETRGTA


```

PERGLYHHHTGDCDKADLSADALRRYRTEOGITLACVFYLAHRSSPILAOIALRLOR
OFDAARVAVGKLVLRPAATYASSESGEDAPKPERVYLIDOLAPILRANADATVAVQAFV
GAMGEWYVYKNGFNAGVATSDOANKRAVYDKLITLTPADRAVQLCTPRKRALYVGG
STRIGHHNDCTFLASDDECTIEDPATREYLYLEETITLPMGSETCAVNPRIKATPAT
AELSRFHWYTLANTYQODVLRAMSAGCLAEVYAKLGRTPLDVSFPTTTRTGALP
VRNLNNDGYAAAVNORTVNLRLNTTGTALTRPLTADPRTMLPGARVIAQOAIQIP
SIATGOYALLLELADPALPGRPREYSIOTANTGIMDARTGINDLKORTVTS"
1825. .2142
/gene="asm36"
1825. .2142
/gene="asm36"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54114.1"
/db_xref="GI:21449382"
/translation="MYTASAVILCARAGCAGPRPAGATARMRECSTACKOAAVRLRKA
GSASHHEDEQLREALLALVARARLTLEATPPWSAQATALVGLARRVVELAARDAGV
SGG"
complement(2129. .3307)
/gene="asm35"
complement(2129. .3307)
/gene="asm35"
/codon_start=1
/transl_table=11
/product="bicyclomycin resistant protein"
/protein_id="AAM54113.1"
/db_xref="GI:21449381"
/translation="MAVGRADSAVGGMLPALLIVALTGTTIEIDMSVPSPFDIARELG
VGGPAVLVTNLFGLYCGALCHGIDSPRGFPVLLGSAVMYLGCALCALAPITG
LLTARFVGGIGSAAAVLVYIGDRYGPALRMVGLVNAAMSALMTCAPLLGGAT
NNVAGRWGYLVACATLALSTLLVFLPETPRARTATARSVADYRLGSRVYL
AASIPRSLLPAAYLVITIASPELTYAFSLGVTGFMHLILVYASFAVVSARPTGITP
LLGGPRGAVRMLGAVYVGVLEFLGIDGPVATGSVALFCGFAVCITYVEDRSKV
FPLLAGAASSATWIGLRALLVTLLTLTLAGALSTGGPPAPAAVMAAVALVLLSLSPR
"
complement(3450. .4064)
/gene="asm34"
complement(3450. .4064)
/gene="asm34"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54112.1"
/db_xref="GI:21449380"
/translation="MAKRNPERPALLDAAVEYLAGEARGLTFFRAVDORAGVPAQTA
SNVFAGRABITGAGERVYORLDDAPLIGETPDGPRDRSVTELMLALVDVRSAPFTG
FLALELRLEAARNPELRVLTARVADIRFNAEHHAAGLPDGTJTTVALMLALMWL
ILERLTPDLTLDHQRRDLVTALVDRLLAGOPPEAGSPDESGK"
4151. .4723
/gene="asm33"
4151. .4723
/gene="asm33"
/codon_start=1
/transl_table=11
/product="dihydrofolate reductase"
/protein_id="AAM54111.1"
/db_xref="GI:21449379"
/translation="MRKLYVFIASIDVYIAGEGHEDEYPTAEPVNEHLRAEYPETT
PTHVRPLVGMPLDTPNRRYDTVMGSGSYOPALDADVGSPIAHROYVVSRSPLDADP
AVALVRADPVGLVRELKREDGLDMLCGGADLAGQLPEIDELVTKTYPVVAGGPTPL
PRAGFAPRAFPRVDVAPRHGGVTTYRPA"
complement(4854. .5057)
/gene="asm32"
complement(4854. .5057)
/gene="asm32"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54110.1"
/db_xref="GI:21449378"
/translation="MWITEONARIPARRSSGSRNARCAVRMPRQPRSATRAPGPRS
SANADSRARSRGDRPTSLNRSR"

```

```

5415. .6461
/gene="asm31"
5415. .6461
/gene="asm31"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor"
/protein_id="AAM54109.1"
/db_xref="GI:21449377"
/translation="MWIGYSADGRRRRRPPVWVGVTAEFTWEDDEDIKLAATLRDREA
ANADSTRAYLKAIGRVRLTLDQEVLELARVAGLYAEHLIDLPPAARSRRPVELR
RVYRDCORADHLEANLRLVYSIAKRYGKGMALLDVIOEGNLIVRAVEKDYAKG
EKSTAYATWMIROAIRALADOSRTIRLPVHWEVYTKIARIERDLHNRGRATAEB
VAEILDVPEKVELROHARTPTSIDOTIGAENDTSIGDLIDENAVADLEAVSSALL
SADLRVLETLTPRENGVVRMRIGFPDGLHTLDEGRYVGYTRENIRIOTESRTMAKL
QHPARSGLEGYLD"
complement(6501. .9518)
/gene="asm30"
complement(6501. .9518)
/gene="asm30"
/codon_start=1
/transl_table=11
/product="cytochrome P450"
/protein_id="AAM54108.1"
/db_xref="GI:21449376"
/translation="MYATGTRIPGPKPLPLVGNLIDVLTSDLTDDVDPLDRCHREHG
YVALITPAGORQVYASSHELVARKCSPPSKKAVHRLDEYRDPAGGLTTPARDEPNW
GKAHRLMPAEGPTAMRDHPAALDIAEQMLVWRFRPGDHRIDVADDDTRLTDLDTIA
LCAFGFNSFYRDPRAHPEVDAMVRSILVEGEAEERLPGVQPLVGRNORYRDDATM
NRJADQIVARARALPAGEPDDILEMLTACADVTGERLSARVRYOLATFLIAGHET
TSGLSLEAFARLHPDEVLARKKADAVDVGDRPAFEDLARLDYGOVLRERLHP
TAPAFLEADPERALEIGHATIGAEPLVLMPTLHRDPAVRDDVDVDRPPEARDE
IPACAMPPRGHGRACITGRPALQERTLVALVLOFPDLALADPDRLITKOTLTIKP
DGLVLRAPRADPGATATVETVYVPHVPTHRHGPPLVFTGSGSGSEGLARTIAG
DSGARMAATSVAPLDDAVRALPASGEVYVSSSYNAPPDNAHFVWMLTODGPDLSG
VDYLVGCGNLDMSATYORVPTLIDEMAAGARLRERGATDARDFGDMERWYEP
LMPLELSAEGCEVEIGEIPRRYVESDAADGLDGLASVAVLENELVGGPDAGSRHLE
LRPLDGTSVRYGDTYLSVLQNHDPDLYRRVARRTAERYVYESSAPGIVYVRAL
RDELLITRCVYDLSAPAGAGVYARLARERCPERRALATGTGTLLELIERPFSCHVD
LALALELPAVRTRLYSTSAAEQERAEVALTVSVTGVSIGTISVRPGGRVAVGAS
PESFEPADNTVPVYLIAAGTGIAPFRGLRLARALGEGEPALLIFCRRGELDDL
YAEFPALGDMLEVDRAVSRHPDGEVRHVOHRIQMORDRVRELVDAGARYVLCGDATR
VGPAAVEVIGRICIPGAGMIDALRAGGRVATDVF"
9664. .10245
/gene="asm29"
9664. .10245
/gene="asm29"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AAM54107.1"
/db_xref="GI:21449375"
/translation="MSAAVRREQLVAALRVKMRDGIAAATTRCAICAEADMPGAFHY
CFRSKOEIYALTTLTGINTSIDAMPALSPDLPAVNVTMLAFTATQADDEALYV
LGEIVTLAOREPELRELPAWQHREYVAKVAVHLDRLAEEGFDYSYDRALLAELVYSV
LDGVYASWESHRRPELGELAKSLDGFAMFAALTR"
10393. .10857
/gene="asm28"
10393. .10857
/gene="asm28"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM54106.1"
/db_xref="GI:21449374"
/translation="MTDTTTRHRTPVQNALAVGAVFLVGIAGFVPGTLTNYDGLFE
AGHHSDAALIGVYKVSALNVLHILAFGAAGIAGLSKAPGAGAFILGGGATYAVLML
XGVYVDEHSTANVPVPLAADNMLHLDLANGMIGLGAALGRNAHYAARASDD"
complement(10912. .11937)
/gene="asm27"
complement(10912. .11937)
/gene="asm27"
/codon_start=1

```


Query Match Best Local Similarity 78.8%; Score: 65.4; DB: 1; Length: 82746;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 GCGCGCGCTTTGCGCCGCGCGCGGTGCAGACGTTCTGCGCGAGCGCGAGCGGCTACGC 60
DB 71233 GCGCGCGCTGTCGCCGCGCGCGGTGCAGGCGTTCTCGCGCGAGCGCGAGCGGCGAGCGC 71174
QY 61 GCGCGCGAGCGGTGCGCGCGGTGCTGCTCAAGCGCGCT 99
DB 71173 GTGGGCTGAGGCGGTGCGCGGTGCTGCTGAGCGCGCT 71135

RESULT 15
LOCUS SAU421825
DEFINITION SAU421825 66808 bp, DNA, linear, BCT 16-MAR-2002
Stigmatella aurantiaca ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
ORF1, stgA gene, stb gene, stc gene, stid gene, stle gene, stlf
gene, stlg gene, stih gene, stlj gene, stlk gene, stll gene and
ORF9.

ACCESSION AJ421825
VERSION AJ421825.1 GI:19572309
KEYWORDS acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
kinase; methyl transferase; ORP1, ORF2, ORF3, ORF4, ORF5, ORF6;
ORF7, ORF8, ORF9; polyketide synthase; stgA gene; stla
channel beta chain; ribosome binding factor; stla gene; stla
gene; stlb gene; stlc gene; stld gene; stle gene; stlf
gene; stlg gene; stih gene; stlj gene; stlk gene; stll gene; stlj
gene; stlj protein; stlk gene; stll gene.

SOURCE Stigmatella aurantiaca
ORGANISM Stigmatella aurantiaca
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Cytophacterales; Cytophacterales; Stigmatella.

REFERENCE 1
AUTHORS Gallatiz, N., Sliakowski, B., Kunze, B., Nordstiek, G., Blocker, H.,
Hofle, G. and Muller, R.

TITLE The biosynthesis of the aromatic myxobacterial electron transport
inhibitor stigmatellin is directed by a novel type of modular
polyketide synthase
Online Publication
J. Biol. Chem., 10.1074/jbc.M11738200
2 (bases 1 to 66808)

JOURNAL Direct Submission
REMARK Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
AUTHORS Brunnschweiler, GERMANY
JOURNAL Location/Qualifiers

FEATURES
source
1. 66808
/organism="Stigmatella aurantiaca"
/strain="Sg a15"
/db_xref="taxon:41"
complement(4978..6331)
/note="ORF8"
/codon_start=1
/transl_table=1
/product="cytochrome P450 monooxygenase"
/protein_id="CAD19077.1"
/db_xref="GI:19572310"

CDS
/transl_table=1
/protein_id="CAD19083.1"
/db_xref="GI:19572316"
/translation="MRPRITGAGGAVPILAMARAVPAGAVKOLLLAROPYLAENVA
AMPVHAEGAVATVPAETERTIVYVGAAGGCGGCGGAVAVRDPFAMSLGRELRLRLA
AFARAVGVGAVHGRSLSLPGDIGRLSTHRLQREMERLVPRKGLTGGGNGHFEEL
DRDAEGALMLLHTGSGVGAALAHHLVAOVAVGESGLGRLRTDSSEGAACLADEL

CDS
/transl_table=1
/codon_start=1
/product="protein kinase"
/protein_id="CAD19078.1"
/db_xref="GI:19572311"
/translation="MEELISALPDGSGVNGVWRLEGKAGTGYCAVRAHVGICR
SVALLANTPMDLNERERALELKERHANYVRLGHWGKRGVGGAEHPIYVMGVES
VRLQWNERELRLAREVRLVAQVRALEALEHMGGLHGDYKGENVLSPGHASLMD
FGCGWBSAALPTGILAPGTAKRSPALNHNHREDAHQAOTPADVVALGAT
AYRGTGTPPATEPSPGVGDERDQVKAQPPSPQPLVPLALPOLILRLMDPKER
GRABELAAEMALERAGAEADEPNNHSSVSEITGRPVSVOREWGWPLVALVA
VGLLWSSGLNTGRRVAVASRMTEGTGVAADAAVEALPVSSAAREPEPSGLMDML
KEPLGQRPPCPFOQISTOGCWLGVSSPPCGESYEYEWKACYSVAAPPRRTS
DPO"

CDS
complement(7849..8577)
/note="ORF6"
/codon_start=1
/transl_table=1
/product="methyl transferase"
/protein_id="CAD19079.1"
/db_xref="GI:19572312"

CDS
/transl_table=1
/product="acyl-CoA binding protein"
/protein_id="CAD19080.1"
/db_xref="GI:19572313"
/translation="MSONIVDDPSPEESYSOLARSLKAGAPAPALRALPDKVL
RYVDLGGCGGCMARKNDAAVHGLIDLSKMLTRAKAETSQDAITYOQADLERVL
PEGAPDLAFSSLAHYEELERLATVYRGITLPGGWFYSIEHPIFASRQPEMTGP
EGHRTMPVDSYOLGSKRTTHMLAREVYKHRTLTITNALLGQFTLRHIDEMQPTA
OLSAPELAERDPMPLVSAOR"

CDS
complement(8622..8879)
/note="ORF5"
/codon_start=1
/transl_table=1
/product="acyl-CoA binding protein"
/protein_id="CAD19081.1"
/db_xref="GI:19572314"
/translation="MALEDERSAQRVAKLTPRPSNDRLLELYSLFKQATEGDVQK
RPGILDYKERAKYDAWGRKGLGRESAAQOYVALVERLRG"

CDS
/transl_table=1
/product="potassium channel beta chain"
/protein_id="CAD19082.1"
/db_xref="GI:19572315"

CDS
/transl_table=1
/product="ribosome binding factor"
/protein_id="CAD19083.1"
/db_xref="GI:19572316"
/translation="WTSFKORRRARPRGTSSLSSEADSLPAHLRVOSTLEQEV
RLRGEISDPLLEGVSVTSFETLPDGRNARGLTLPPEYAGGSAPTEALAVRASGFI
RAQALGINTLKRVPRLRVCVAPGSPSIPEDSABEGGS"
10859..12016
/note="ORF2"
/codon_start=1
/transl_table=1
/protein_id="CAD19084.1"
/db_xref="GI:19572317"

CDS
/transl_table=1
/protein_id="CAD19085.1"
/db_xref="GI:19572318"
/translation="MRPRITGAGGAVPILAMARAVPAGAVKOLLLAROPYLAENVA
AMPVHAEGAVATVPAETERTIVYVGAAGGCGGCGGAVAVRDPFAMSLGRELRLRLA
AFARAVGVGAVHGRSLSLPGDIGRLSTHRLQREMERLVPRKGLTGGGNGHFEEL
DRDAEGALMLLHTGSGVGAALAHHLVAOVAVGESGLGRLRTDSSEGAACLADEL

[illegible]

PA	(NOVS) NOVARTIS-ERRINDUNGEN VERN GSS MBH.
XX	
XX	Schupp T, Ligon JM, Molnar I, Zikkle R, Goerlach J, Cyr D:
XX	WPI: 2000-007744/08
DR	P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58578,
DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR	AAY58592, AAY58593, AAY58594.
XX	
PT	New isolated epoethione synthase genes, used for the recombinant
PT	production of epoethione for use in cancer therapy
XX	
XX	Claim 14; Page 87-104; 174pp: English.
XX	
CC	This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC	comprising 22 open reading frames (ORFs) and includes genes encoding
CC	proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC	starter unit; polyketides being synthesised from two-carbon building
CC	blocks, the beta-carbon of which always carries a keto group. Each round
CC	of two-carbon addition is carried out by a complex of enzymes known as
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC	EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC	the thiazole ring formation of epoethiones, and EPOS C, EPOS D
CC	and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC	formation. EPO F (AAY58582) and Orf14 (AAY58593) are thought to be
CC	the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC	involved in transport. Epoethiones mimic the biological activity of
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic
CC	compositions. Epoethiones exhibit a much lower drop in potency against a
CC	multiply drug-resistant cell line compared with taxol, and are
CC	considerably less efficiently exported from such cells by the multidrug
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC	epoethiones as anticancer agents, they are problematical to produce on a
CC	large scale. Epoethiones are too complex for industrial scale chemical
CC	synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC	poor yields of epoethiones. The nucleic acids of the invention may be
CC	used for the recombinant production of epoethiones in a heterologous host
CC	that is more amenable to fermentation.
XX	
SO	Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other:
XX	
Query Match	100.0%; Score 101; DB 21; Length 68750;
Best Local Similarity	100.0%; Prid. No. 5, 6e-16;
Matches 101; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 GCGCGCGCTTCGCCCGCAGCGGCGGTGCAAGACGTTCGCGCGCAGCGAGCGACTACGC 60
DB	31462 GCGCGCGCTTCGCCCGCAGCGGCGGTGCAAGACGTTCGCGCGCAGCGAGCGACTACGC 31521
OY	61 GCGGGGCGAGGGGCGTGCAGCGGTGATGCTCTCAAGCGGCTGC 101
DB	31522 GCGGGGCGAGGGGCGTGCAGCGGTGATGCTCTCAAGCGGCTGC 31562
RESULT 2	
ID	AAA29349
AC	AAA29349; DNA; 71989 BP.
XX	
XX	12-SEP-2000 (first entry)
XX	
DE	Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX	
KM	Epoethione: polyketide synthase: epocA; epob; epoc; epod; epoE; epof;
KM	epol; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW	tubulin polymerization assay; anti-tumour; cyostatic; ds.
XX	
OS	Sorangium cellulosum.
XX	
Key	Location/Qualifiers

```
FT CDS 3..992
FT /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKs"
FT 989..1501
FT /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKs"
FT 1998..6263
FT CDS
FT /*tag= c
FT /label= epOA_gene
FT /note= "encodes the loading domain"
FT 2031..3548
FT /*tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661
FT /*tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5810
FT /*tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6135
FT /*tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493
FT /*tag= h
FT /label= epOB_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548
FT /*tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /*tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887
FT /*tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887
FT /*tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366
FT /*tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921
FT /*tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921
FT /*tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308
FT /*tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422
FT /*tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905
FT /*tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983
FT /*tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /*tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /*tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /*tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306
FT /*tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137
FT /*tag= x
FT /label= epOC_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /*tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /*tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /*tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756
FT /*tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008
FT /*tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /*tag= ad
FT /label= epOD_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /*tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /*tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /*tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /*tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /*tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /*tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647
FT /*tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /*tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /*tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /*tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574
FT /*tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287
FT /*tag= ap
```

```
FT      misc_RNA      /note= "encodes IRS"
FT      30321..30869
FT      /tag= aq
FT      /note= "encodes IRS"
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes PCPS"
FT      31440..32807
FT      /tag= as
FT      /note= "encodes IRS6"
FT      33018..34067
FT      /tag= at
FT      /note= "encodes A16"
FT      34107..34676
FT      /tag= au
FT      /note= "encodes E36"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes E16"
FT      36705..37256
FT      /tag= aw
FT      /note= "encodes K16"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes A1P6"
FT      37912..49308
FT      /tag= ay
FT      /label= epoc gene
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes K57"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes A17"
FT      41341..41922
FT      /tag= bb
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score: 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred No. 5,6e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GCGCGCGCTTTGCGCCGACGCGGCGGTGCAAGACGTTCTCGCGCGACGCGGCTACGC 60
DB      25850 GCGCGCGCTTTGCGCCGACGCGGCGGTGCAAGACGTTCTCGCGCGACGCGGCTACGC 25909
OY      61 GCGGCGGCGAGGGGTGCGCGGTGTGTGCTCAAGCGGCTGC 101
DB      25910 GCGGCGGCGAGGGGTGCGCGGTGTGTGCTCAAGCGGCTGC 25950

RESULT 3
AAS17367
ID      AAS17367 standard; DNA; 33529 BP.
AC      AAS17367;
XX
XX
DT      12-MAR-2002 (first entry)
XX
XX
DE      DNA sequence of S. cellulosum polyketide synthase cosmid, PKOS28-26.
XX
XX
KW      Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW      acyl transferase domain; dehydratase domain; ketoreductase domain;
KW      acyl carrier protein domain; PKOS28-26; ds.
XX
XX
OS      Sorangium cellulosum.
XX
XX
PN      US6280999-B1.
XX
PD      28-AUG-2001.
```

```
PF      31-AUG-1998; 98US-0144085.
XX
XX      22-JAN-1998; 98US-0010809.
XX
XX      (KOSA-) KOSAN BIOSCIENCE.
XX
XX      Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX      WPI; 2001-606536/69.
XX
XX      Novel purified, isolated DNA molecule from Sorangium cellulosum having
XX      polyketide open reading frame encoding modules with one or more domains
XX      such as Ketosynthase, acyl transferase and acyl carrier protein domains
XX
XX      Claim 4; Fig 1; 72pp; English.
XX
XX      The present invention relates to the isolation of novel Sorangium
XX      cellulosum polyketide synthases (PKS), and the polynucleotide sequences
XX      encoding them. The polyketide synthases include catalytic domains such
XX      as ketosynthase domain, acyl transferase domain, dehydratase domain,
XX      ketoreductase domain and acyl carrier protein domain. A host cell
XX      comprising a PKS ORF (open reading frame) which encodes one or more
XX      more PKS domains is useful for producing polyketide synthases from which
XX      polyketides can be produced. The host cells are useful for constructing
XX      a library, where each individual colony of the library represents a
XX      colony with the ability to produce a particular PKS synthase and
XX      ultimately a particular polyketide. The polyketides produced by these
XX      colonies can be used collectively in a panel to represent a library or
XX      may be assessed individually for activity. Colonies in the library are
XX      also induced to produce the relevant synthases and thus to produce the
XX      relevant polyketides to obtain a library of candidate polyketides which
XX      can be screened for binding to desired targets such as receptors
XX      signalling proteins, etc. The present sequence represents the DNA
XX      sequence of cosmid PKOS28-26 which encodes one or more domains of
XX      S. cellulosum PKS.
XX
XX      Note: The present sequence is said to encode the functional domains
XX      of S. cellulosum PKS which correspond to domains or domain subsets of
XX      the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ      Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

Query Match      64.8%; Score 65.4; DB 23; Length 33529;
Best Local Similarity 78.8%; Pred. No. 2.1e-07;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY      1 GCGCGCGCTTTGCGCCGACGCGGCGGTGCAAGACGTTCTCGCGCGACGCGGCTACGC 60
DB      10351 GCGTGGGTCTCGCCCGACGCGGCGGTGCAAGACGTTCTCGCAAGCGCGGCTACGC 10410
OY      61 GCGGCGGCGAGGGGTGCGCGGTGTGTGCTCAAGCGGCT 99
DB      10411 CCGCGGCGAGGGGTGCGCGGTGTGTGCTCAAGCGGCT 10449

RESULT 4
AAF89974
ID      AAF89974 standard; DNA; 764 BP.
AC      AAF89974;
XX
XX
DT      06-AUG-2001 (first entry)
XX
XX
DE      Partial nucleotide sequence of a type I polyketide synthase.
XX
XX      Metabolic pathway operon; polyketide; polyketide antibiotic;
XX      type I polyketide synthase; ss.
XX
XX      Unidentified.
XX
XX
PN      WO200140497-A2.
XX
PD      07-JUN-2001.
```

```
XX 27-NOV-2000; 2000WO-FR0311.
PE 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET ) AVENTIS PHARMA SA.
XX
PI Jeanin P, Perodet J, Guertineau M, Simonet P, Courtois S,
PI Cappellano C, Francou F, Raynal A, Ball M, Sezouov G, Taphile K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
P-PSDB: AAB83966.
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polypeptide synthases and derived
PT antibiotics
XX
PS Claim 35; Page 224; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticles;
CC passing nucleic acid-containing solution through a molecular sieve;
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polypeptide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polypeptide antibiotics. AAF89964-78
CC encode partial type I polypeptide synthases, and were isolated using the
CC method of the invention.
XX
SQ Sequence 764 BP; 114 A; 272 C; 268 G; 110 T; 0 other;
XX
Query Match 62.0%; Score 62.6; DB 22; Length 764;
Best Local Similarity 79.6%; Pred. No. 1.2e-06;
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
XX
QY 7 GCTTTCGCCACGCGGCTCTCGGCGGACGCGGACGCGCTACGCGCGG 66
Db 480 GCTGCGCGCTACGCGGCGCTCGACGCTTTCGACGCGCGGCGGACGCGCTACGCGCGG 539
XX
QY 67 CGAGGGGTGCGCGCTGTGCTCAAGCGGCT 99
Db 540 CGAAGGGCGCGGCTGATCGTCAAGCGGCT 572
XX
RESULT 5
AA19683/C
ID AA19683 standard; DNA; 6564 BP.
XX
AC AA19683;
XX
DT 09-APR-2001 (first entry)
XX
DE Mycobacterium tuberculosis ppsc gene.
XX
KW Mycobacterium tuberculosis; attenuated microorganism;
KW signature tagged transposon mutant; mutant library;
KW mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW vaccine; ppsc; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200102555-A1.
XX
PD 11-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-IB00950.
XX
```

```
PR 06-JUL-1999; 99US-0142982.
PR 08-JUL-1999; 99US-0142833.
XX
PA (INSP ) INST PASTEUR.
XX
PI Glacquel B, Guilhot C, Camacho L;
XX
DR WPI: 2001-091804/10.
P-PSDB: AAB66467.
XX
XX Screening a mutant library for mutants unable to grow under specific
PT conditions and for identifying loci involved in pathogenicity,
PT comprises using signature tagged transposon mutagenesis
XX
PS Example 8; Page 122-124; 159pp; English.
XX
XX The present sequence is given in a specification relating to a method for
CC screening a library of mutants. The method comprises constructing a
CC library with insertions in genes and/or regulatory regions of the
CC organisms of interest, where the insertion contains a tag and/or a
CC transposon associated with a tag. The mutants are identified by
CC hybridisation of the tags to known sequences. The method is useful for
CC treating an individual suffering from a mycobacterial infection,
CC suspected of being infected with a Mycobacterium, or having been
CC exposed to an infectious Mycobacterium. It is also useful for
CC identifying and isolating mutants of actinomycetales and for identifying
CC compounds that have antibiotic activity. The method is used to identify
CC mutants of microorganisms, preferably an actinomycetales, such as
CC M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
CC M. paratuberculosis, that is unable to grow under specific conditions.
CC It is especially useful for identifying loci involved in pathogenicity.
CC It is useful in constructing vaccines. The method can be used to screen
CC multiple libraries concurrently. It can screen libraries of different
CC organisms or different strains of the same organism.
XX
SQ Sequence 6564 BP; 1137 A; 2189 C; 2154 G; 1084 T; 0 other;
XX
Query Match 61.4%; Score 62; DB 22; Length 6564;
Best Local Similarity 78.7%; Pred. No. 1.5e-06;
Matches 74; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
XX
QY 7 GCTTTCGCCACGCGGCGGCTCTCGGCGGACGCGGACGCGCTACGCGCGG 66
Db 762 GCTGCGCGCTACGCGGCGGCTCGACGCTTTCGACGCGCGGCGGACGCGCTACGCGCGG 821
XX
QY 67 CGAGGGGTGCGCGCTGTGCTCAAGCGGCTG 100
Db 822 CGAAGGGCGCGGCTGATCGTCAAGCGGCTG 855
XX
RESULT 6
AA199683/C
ID AA199683 standard; DNA; 4403765 BP.
XX
AC AA199683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
```

PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraher CM, Venter JC;
XX
XX WPI: 2001-647261/74.
DR
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PI genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ.
XX
PS Claim 4: SEQ ID NO 2: 3pp + Sequence Listing: English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
Query Match 60.8%; Score 61.4; DB 22: Length 4403765;
Best Local Similarity 77.9%; Pred. No. 1.4e-06;
Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
OY 5 GCGCTTCCGCGGCGGCGGTCGCAAGAGCTTCGCGCGGCGGCGGCGGCGGCGG 64
DB 4252406 GTGCTGGCGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4252347
OY 65 GCGGAGGGGTGCGCGCTGCTGCTCAAGCGGCT 99
DB 4252346 TCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4252312
RESULT 7
AA199682/C
ID AA199682 standard; DNA; 4411529 BP.
XX
XX AA199682;
AC
XX
XX 15-JAN-2002 (first entry)
DT
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
DE
XX
XX Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX US6294328-B1.
PN
XX
XX 25-SEP-2001.
PD
XX
XX 24-JUN-1998; 98US-0103840.
PF
XX
XX 24-JUN-1998; 98US-0103840.
PR
XX
XX (GENO-) INST GENOMIC RES.
PA
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
PI
XX
XX WPI: 2001-647261/74.
DR
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the

PT genome corresponding to positions where M. tuberculosis strains CDC
PI 1551 and H37Rv differ.
XX
XX
PS Claim 3: SEQ ID NO 1: 3pp + Sequence Listing: English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 144602 G; 758379 T; 0 other;
Query Match 60.8%; Score 61.4; DB 22: Length 4411529;
Best Local Similarity 77.9%; Pred. No. 1.4e-06;
Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
OY 5 GCGCTTCCGCGGCGGCGGTCGCAAGAGCTTCGCGCGGCGGCGGCGGCGGCGGCGG 64
DB 4260150 GTGCTGGCGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4260091
OY 65 GCGGAGGGGTGCGCGCTGCTGCTCAAGCGGCT 99
DB 4260090 TCCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4260056
RESULT 8
AA199970
ID AA199970 standard; DNA; 758 BP.
XX
XX AA199970;
AC
XX
XX 06-AUG-2001 (first entry)
DT
XX
XX Partial nucleotide sequence of a type I polyketide synthase.
DE
XX
XX Metabolic pathway operon: polyketide; polyketide antibiotic;
KW type I polyketide synthase; ss.
XX
XX Unidentified.
OS
XX
XX WO200140497-A2.
PN
XX
XX 07-JUN-2001.
PD
XX
XX 27-NOV-2000; 2000WO-FR03311.
PF
XX
XX 29-NOV-1998; 99FR-0015032.
PR
XX
XX 07-JUN-2000; 2000US-0209800.
PI
XX
XX (AVENTIS PHARMA SA.
PA
XX
XX Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francon F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
DR
XX
XX P-PSDB; AAB83962.
PD
XX
XX Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics.
XX
PS Claim 35; Page 222; 356pp; French.

CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polypeptide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polypeptide antibiotics. AAF8964-78
CC encode partial type I polypeptide synthetases, and were isolated using the
CC method of the invention.

CC
XX
SQ Sequence 758 BP; 121 A; 250 C; 247 G; 140 T; 0 other;

Query Match 60.6%; Score 61.2; DB 22; Length 758;
Best Local Similarity 76.5%; Pred. No. 2.7e-06;
Matches 75; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 2 CGCGGCGTTTCCGCCGAGCGGGGTGCAAGAGCTTCTGGCCGAGCGGCGTACGCG 61
DB 469 CGCATGTTGGCGCCCGACGAGCGCTGCAAGAGCTTCTGAGCGCGAGCGGTTATGTG 528
OY 62 CGGCGGAGAGGGGTGCGCGCTGTGTGCTCAAGCGGCT 99
DB 529 CGCGGCGAGGGGTGCGCGCTGTGTGCTGTGCAAGCGGCT 566

RESULT 9
AAF90038
ID AAF90038 standard; DNA; 5088 BP.
XX
AC AAF90038;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polypeptide synthase.
XX
KW Metabolic pathway operon; polypeptide; polypeptide antibiotic;
XX type I polypeptide synthase; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT 1..5088
FT CDS /tag= a
FT /product= "type I polypeptide synthase"
FT /transl_except= "1..3, aa: Met"

XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PE 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
XX 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
XX P-PSDB; AAB83975.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polypeptide synthetases and derived
XX antibiotics
XX
PS Claim 35; Page 309-311; 356pp; French.

XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polypeptide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polypeptide antibiotics. AAF9034-39
CC represent open reading frames (ORFs) of the coding strand of cosmid
CC a26g1, and encode type I polypeptide synthetases.

CC
XX
SQ Sequence 5088 BP; 838 A; 1622 C; 1679 G; 949 T; 0 other;

Query Match 60.6%; Score 61.2; DB 22; Length 5088;
Best Local Similarity 76.5%; Pred. No. 2.4e-06;
Matches 75; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 2 CGCGGCGTTTCCGCCGAGCGGGGTGCAAGAGCTTCTGGCCGAGCGGCGTACGCG 61
DB 739 CGCATGTTGGCGCCCGACGAGCGCTGCAAGAGCTTCTGAGCGCGAGCGGTTATGTG 798
OY 62 CGGCGGAGAGGGGTGCGCGCTGTGTGCTCAAGCGGCT 99
DB 799 CGCGGCGAGGGGTGCGCGCTGTGTGCTGTGCAAGCGGCT 836

RESULT 10
AAF90033
ID AAF90033 standard; DNA; 34071 BP.
XX
AC AAF90033;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KW Metabolic pathway operon; polypeptide; polypeptide antibiotic; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 1..34071
FT CDS /tag= a
FT /product= "type I polypeptide synthase"
FT /transl_except= "1..3, aa: Met"

XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PE 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
XX 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
XX WPI: 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polypeptide synthetases and derived
XX antibiotics
XX
XX Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX passing nucleic acid-enriched fractions through an anion exchange


```

Matches      74 Conservative    0 Mismatches   22 Indels     0 Gaps
OY          5 GGCGTTTGCCCGACGGGCGGTGCAGAACGCTTCTGGGCCGACGCGAGCGGTACGCGC 64
              ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db         182 GGCGTGTACCACGCACCGCGCATGTCATCAAGCTTCGACGCGCACGCGATGAATTCCGTCGC 123
               ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++
OY          65 GGCGAGGGGTGGCGCGGTGGTGGTCMAAGCGGCG 100
              ||||||| | | | | | | | | | | | | | | | | | | | | | | |
Db        122 GGCGAGGGGTGGCGGTGGTGGTCTCAAGCGCTTG 87
                ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++

RESULT_13
ID AAF31638 standard; DNA; 5481 BP.
AAF31638
XX AAF31638;
AC
XX
XX AAFA31638;
AC
XX
XX 09-APR-2001 (first entry)
DT
DX Mycobacterium tuberculosis ppsd gene.
DE
KW Mycobacterium tuberculosis; attenuated microorganism;
RV signature tagged transposon mutant; mutant library;
RM mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
RW vaccine; ppsd; ds.
KM
XX Mycobacterium tuberculosis.
OS
PN WO200102555-A1.
PD
PX 11-JAN-2001.
PF
PP 06-JUL-2000; 2000WO-IB00950.
PR
PS 06-JUL-1999; 99US-O142982.
PT
PZ 08-JUL-1999; 99US-O142833.
XA
XN (INSP ) INST PASTEUR.
YP
PI Glacquel B, Guilhot C, Camacho L;
DR WPl: 2001-091804/10.
DS P-PsDB; AAB66468.
XX
XX Screening a mutant library for mutants unable to grow under specific
PT conditions and for identifying loci involved in pathogenicity,
PT comprises using signature tagged transposon mutagenesis -
XX
XX Example 8; Page 131-132; 159pp; English.
FX
XX The present sequence is given in a specification relating to a method for
SC screening a library of mutants. The method comprises constructing a
CC library with insertions in genes and/or regulatory regions of the
CC organisms of interest, where the insertion contains a tag and/or a
CC transposon associated with a tag. The mutants are identified by
CC hybridisation of the tags to known sequences. The method is useful for
CC treating an individual suffering from a mycobacterial infection,
CC suspected of being infected with a mycobacterium, or having been
CC exposed to an infectious Mycobacterium. It is also useful for
CC identifying and isolating mutants of actinomyces and for identifying
CC compounds that have antibiotic activity. The method is used to identify
CC mutants of microorganisms, preferably an actinomycetes, such as
CC M. tuberculosis, M. boyds, M. leprae, M. avium, M. intracellulare and
CC M. paratuberculosis, that is unable to grow under specific conditions.
CC It is especially useful for identifying loci involved in pathogenesis.
CC It is useful in constructing vaccines. The method can be used to screen
CC multiple libraries concurrently. It can screen libraries of different
CC organisms or different strains of the same organism.
CX
XX Sequence 5481 BP; 894 A; 1745 C; 1925 G; 917 T; 0 other;
```

```

Matches 74: Conservative 0; Mismatches 22; Indels 0; Gaps 0
OY 5 GCGCTTTCGCCCGACGGCGGTGCAAGACGTTCTCGCGCGACGCGAGCGGCTACGCGCG 64
    |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 751 GCGGTGTACACCGACGCGCGGATGCAACAGCTTCGACGCCAACGCGGATGATTCGTGCGC 810
OY 65 GCGGAGGGGTGCGCGGTGTGTGCTGTCAAGCGCGCG 100
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 811 GCGGAGGCGCTGCGCGGTGTGTGCTGTCAAGCGGTTG 846

RESULT 14
AAH52061
ID AAH52061 standard; DNA; 5485 BP.
XX
XX AAH52061;
AC
XX
DT 04-SEP-2001 (first entry)
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 115.
XX
XX Drug target; growth; organism viability; characterisation; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX Mycobacterium tuberculosis.
XX WO200135317-A1.
XX
XX 17-MAY-2001.
XX
XX 13-NOV-2000; 2000WO-US31152.
XX
XX 12-NOV-1999; 99US-0165086.
XX 12-NOV-1999; 99US-0165124.
XX 01-FEB-2000; 2000US-0179531.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Eisenberg D, Rotstein SH, Marcotte EM;
PI
DR WPI: 2001-329193/34.
XX
XX P-PSDB; AAG81210.
XX
XX
XX Identifying nucleotide or polypeptide sequence for use as drug target,
XX involves providing algorithm that analyzes a functional relationship
XX between nucleotide or polypeptide sequences, and comparing the
XX sequences
XX
XX Disclosure: Page 128-130; 207pp; English.
XX
XX This invention relates to a method for identifying a nucleotide or
XX polypeptide sequence that may be a drug target, or essential for growth
XX or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
XX represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
XX tuberculosis proteins which are potential drug targets. The DNA and
XX protein sequences are used to illustrate the method of the invention. The
XX method involves providing an unknown nucleotide or polypeptide sequences,
XX and comparing it to a number of sequences along with at least one
XX algorithm capable of analysing a functional relationship between
XX nucleotide and polypeptide sequences. The method is useful for
XX characterising the function of nucleic acids and polypeptides that may be
XX useful as a target for a drug or essential for the growth or viability of
XX an organism.
XX
XX Sequence 5485 BP; 896 A; 1745 C; 1926 G; 918 T; 0 other;
XX
XX
XX Query Match 60.2%; Score 60.8; DB 22; Length 5485;
XX Best Local Similarity 77.1%; Pred. No. 3e-06;
XX Matches 74; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
OY 5 GCGCTTTCGCCCGACGGCGGTGCAAGACGTTCTCGCGCGACGCGGAGCGGCTACGCGCG 64
    |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 751 GCGGTGTACACCGACGCGCGGATGCAACAGCTTCGACGCCAACGCGGATGATTCGTGCGC 810
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

OY 65 GCGAGGGTGGCCGCTGCTGCTCAGCGCTG 100
||||| ||| ||||||||| |||
DB 811 GCGAGGCTGGCGCTGCTGCTCAGCGCTTG 846

RESULT 15

AAS08693
ID AAS08693 standard: DNA; 109519 bp.

AC AAS08693;

DT 26-SEP-2001 (first entry)

XX Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.

XX Everninomycin; antibiotic; bottle-neck gene; orthomycin.

KW Fermentation; ds.

XX Micromonospora carbonacea var. africana.

XX Location/Qualifiers

FT Key complement (132..1382)

FT CDS /tag= a

FT RBS /product= "Evda"

FT CDS /tag= b

FT RBS /product= "Evdb"

FT CDS /tag= c

FT RBS /product= "Evdc"

FT CDS /tag= d

FT RBS /product= "Evde"

FT CDS /tag= e

FT RBS /product= "Evdf"

FT CDS /tag= f

FT RBS /product= "Evdg"

FT CDS /tag= g

FT RBS /product= "Evdh"

FT CDS /tag= h

FT RBS /product= "Evdh"

FT CDS /tag= i

FT RBS /product= "Evid"

FT CDS /tag= j

FT RBS /product= "Evdj"

FT CDS /tag= k

FT RBS /product= "Evdj"

FT CDS /tag= l

FT RBS /product= "Evdg"

FT CDS /tag= m

FT RBS /product= "Evdh"

FT CDS /tag= n

FT RBS /product= "Evdh"

FT CDS /tag= o

FT RBS /product= "Evdj"

FT RBS complement (13027..13030)

FT CDS /tag= t

FT RBS /product= "Evda"

FT CDS /tag= u

FT RBS /product= "Evda"

FT CDS /tag= v

FT RBS /product= "Evdb"

FT CDS /tag= w

FT RBS /product= "Evdb"

FT CDS /tag= x

FT RBS /product= "Evdc"

FT CDS /tag= y

FT RBS /product= "Evdc"

FT CDS /tag= z

FT RBS /product= "Evde"

FT CDS /tag= aa

FT RBS /product= "Evde"

FT CDS /tag= ab

FT RBS /product= "Evdf"

FT CDS /tag= ac

FT RBS /product= "Evdf"

FT CDS /tag= ad

FT RBS /product= "Evde"

FT CDS /tag= ae

FT RBS /product= "Evde"

FT CDS /tag= af

FT RBS /product= "Evdf"

FT CDS /tag= ag

FT RBS /product= "Evdf"

FT CDS /tag= ah

FT RBS /product= "Evde"

FT CDS /tag= ai

FT RBS /product= "Evde"

FT CDS /tag= aj

FT RBS /product= "Evde"

FT CDS /tag= ak

FT RBS /product= "Evde"

FT CDS /tag= al

FT RBS /product= "Evde"

FT CDS /tag= am

FT RBS /product= "Evde"

FT RBS complement (14410..15363)

FT CDS /tag= u

FT RBS /product= "Evda"

FT CDS /tag= v

FT RBS /product= "Evdb"

FT CDS /tag= w

FT RBS /product= "Evdb"

FT CDS /tag= x

FT RBS /product= "Evdc"

FT CDS /tag= y

FT RBS /product= "Evdc"

FT CDS /tag= z

FT RBS /product= "Evde"

FT CDS /tag= aa

FT RBS /product= "Evde"

FT CDS /tag= ab

FT RBS /product= "Evdf"

FT CDS /tag= ac

FT RBS /product= "Evdf"

FT CDS /tag= ad

FT RBS /product= "Evde"

FT CDS /tag= ae

FT RBS /product= "Evde"

FT CDS /tag= af

FT RBS /product= "Evdf"

FT CDS /tag= ag

FT RBS /product= "Evdf"

FT CDS /tag= ah

FT RBS /product= "Evde"

FT CDS /tag= ai

FT RBS /product= "Evde"

FT CDS /tag= aj

FT RBS /product= "Evde"

FT CDS /tag= ak

FT RBS /product= "Evde"

FT CDS /tag= al

FT RBS /product= "Evde"

FT CDS /tag= am

FT RBS /product= "Evde"

FT CDS /tag= an

FT RBS /product= "Evde"

FT RBS complement (34449..35210)

FT CDS /tag= ao

FT RBS /product= "Evde"

FT CDS /tag= ap

FT RBS /product= "Evde"

FT CDS /tag= aq

FT RBS /product= "Evde"

FT CDS /tag= ar

FT RBS /product= "Evde"

FT CDS /tag= as

FT RBS /product= "Evde"

FT CDS /tag= at

FT RBS /product= "Evde"

FT CDS /tag= au

FT RBS /product= "Evde"

FT CDS /tag= av

FT RBS /product= "Evde"

FT CDS /tag= aw

FT RBS /product= "Evde"

FT CDS /tag= ax

FT RBS /product= "Evde"

FT CDS /tag= ay

FT RBS /product= "Evde"

FT CDS /tag= az

FT RBS /product= "Evde"

FT CDS /tag= ba

FT RBS /product= "Evde"

FT CDS /tag= bb

FT RBS /product= "Evde"

FT CDS /tag= bc

FT RBS /product= "Evde"

FT CDS /tag= bd

FT RBS /product= "Evde"

FT CDS /tag= be

FT RBS /product= "Evde"

FT CDS /tag= bf

FT RBS /product= "Evde"

FT CDS /tag= bg

FT RBS /product= "Evde"

FT CDS /tag= bh

FT RBS /product= "Evde"

Search completed: November 5, 2002, 13:36:41
Job time : 1694.22 secs

```
FT      /*tag- au
FT      /product- "EvrS"
FT      complement (40216..40890)
FT      /*tag- av
FT      /product- "EvrT"
FT      complement (40899..40902)
FT      /*tag- aw
FT      complement (40887..41576)
FT      /*tag- ax
FT      /product- "EvrU"
FT      complement (41679..42707)
FT      /*tag- ay
FT      /product- "EvrV"
FT      complement (42714..42717)
FT      /*tag- az
FT      complement (42810..43799)
FT      /*tag- Da
FT      /product- "EvrW"
FT      complement (43807..43811)
FT      /*tag- Db
FT      complement (43799..44866)
FT      /*tag- Dc
FT      /product- "EvrX"
FT      complement (45014..45760)
FT      /*tag- Dd
FT      /product- "EvrY"
FT      complement (45767..45770)
FT      /*tag- De
FT      complement (45962..46714)
FT      /*tag- Df
FT      /product- "EvrZ"
FT      complement (45952..45956)
FT      /*tag- Dg
FT      complement (47156..49234)
FT      /*tag- Dh
FT      /product- "EvrA"
FT      complement (51627..52715)
FT      /*tag- Di
FT      /product- "EvrB"
FT      complement (51629..51622)
FT      /*tag- Dj
FT      complement (52889..53557)
FT      /*tag- Dk
FT      /product- "EvrC"
FT      complement (53554..54207)
FT      /*tag- Dl
FT      /product- "EvrD"
FT      complement (54362..55117)
FT      /*tag- Dm
FT      /product- "EvrB"
FT      complement (55125..55128)
FT      /*tag- Dn
FT      complement (55135..56094)
FT      /*tag- Do
FT      /product- "EvrC"
FT      complement (56100..56103)
FT      /*tag- Dp
FT      complement (56184..56813)
FT      /*tag- Dq
FT      /product- "EvrC2"
FT      complement (56961..58709)
FT      CDS
```

Query Match 59.6%; Score 60.2; DB 22; Length 109519;
Best Local Similarity 76.3%; Pred. No. 3,4e-06;
Matches 74; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```
OY      3 GCGGCTTTCGCCGAGCGGCGGCTTCGCGCGAGCGGCGGCTACGCGC 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 27367 GTGCCCTGGCAGCCGAGGCGGCGAGCAAGTCCTTCGAGCGCCGCGGCTACGCTC 27426

OY      63 GGGGCGAGGGGTGCGCGTGTGTGTGCTCAAGCGGCT 99
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 27427 GTGGCGAGGGGTGCGCGTGTGTGCTCAAGCTGCT 27463
```


COMMITTEE ON INFORMATION: 03/03/2003 FOR

```
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 60
    |||||||
DB 31462 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 31521
    |||||||
QY 61 GCGGGGCGAGGGGTGCGCCGCGTGGTGTCAGCGGCGCTGC 101
    |||||||
DB 31522 GCGGGGCGAGGGGTGCGCCGCGTGGTGTCAGCGGCGCTGC 31562
    |||||||
```

```
RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 635457
```

```
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 60
    |||||||
DB 31462 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 31521
    |||||||
QY 61 GCGGGGCGAGGGGTGCGCCGCGTGGTGCTAAGCGGCTGC 101
    |||||||
DB 31522 GCGGGGCGAGGGGTGCGCCGCGTGGTGCTAAGCGGCTGC 31562
    |||||||
```

```
RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
```

```
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
```

```
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 60
    |||||||
DB 31462 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 31521
    |||||||
QY 61 GCGGGGCGAGGGGTGCGCCGCGTGGTGCTCAAGCGGCTGC 101
    |||||||
DB 31522 GCGGGGCGAGGGGTGCGCCGCGTGGTGCTCAAGCGGCTGC 31562
    |||||||
```

```
RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 635459
```

```
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1
```

```
Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 60
    |||||||
DB 31462 GCGCGGCGCTTTGCGCCCGACGGGGGCGGTGCAAGACGTTCTCGGGCCGACGGCGGCTACGC 31521
    |||||||
QY 61 GCGGGGCGAGGGGTGCGCCGCGTGGTGCTCAAGCGGCTGC 101
    |||||||
DB 31522 GCGGGGCGAGGGGTGCGCCGCGTGGTGCTCAAGCGGCTGC 31562
    |||||||
```

```
RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
```

```
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
```



```
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGACGTTCTCGGCGGACGCGGCTACGC 60
Db 31462 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGACGTTCTCGGCGGACGCGGCTACGC 31521
QY 61 GCGGCGGAGGGGTGCGCCGCTGTGTGCTCAAGCGGCTGC 101
Db 31522 GCGGCGGAGGGGTGCGCCGCTGTGTGCTCAAGCGGCTGC 31562

RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGACGTTCTCGGCGGACGCGGCTACGC 60
Db 31462 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGACGTTCTCGGCGGACGCGGCTACGC 31521
QY 61 GCGGCGGAGGGGTGCGCCGCTGTGTGCTCAAGCGGCTGC 101
Db 31522 GCGGCGGAGGGGTGCGCCGCTGTGTGCTCAAGCGGCTGC 31562

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:

; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match          100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGACGTTCTCGGCGGACGCGGCTACGC 60
Db 25850 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGACGTTCTCGGCGGACGCGGCTACGC 25909
QY 61 GCGGCGGAGGGGTGCGCCGCTGTGTGCTCAAGCGGCTGC 101
Db 25910 GCGGCGGAGGGGTGCGCCGCTGTGTGCTCAAGCGGCTGC 25950

RESULT 9
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYPEPTIDE SYNTHASES AND ENCODING DNA
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3

Query Match          64.8%; Score 65.4; DB 4; Length 33529;
Best Local Similarity 78.8%; Pred. No. 2.4e-08;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGACGTTCTCGGCGGACGCGGCTACGC 60
||| ||| | ||||||||| ||||||||| ||||||||| | |||||||||
```

Db 10351 GCGTGGGCTCTGCGCCGAGCGGTCAGACGTTCTGCGAGCGCGCGGCTACCG 10410
QY 61 GCGGGGCGAGGGGTGCGCGGTGGTG-TCACGCGCT 99
Db 10411 CCGCGGCGAGGCGTGGGGTGTGTGTG-TCATGCGCT 10449

RESULT 10
US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 60.8%; Score 61.4; DB 4; Length 4403765;
Best Local Similarity 77.9%; Pred. No. 1.8e-07;
Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 5 GGGCTTGGCCGAGGGGGGTGCAAGACGTTCTGCGCGGAGCGGCTACGCGCGG 64
Db 4252406 GTGCTGGCGCCCGAGCGCGCGGATCAAGTCTTCTACGCGAGCGCGGCTACACCGC 4252347
QY 65 GCGGAGGGGTGCGCGGTGTGCTCTCAAGCGGCT 99
Db 4252346 TCCGAGGCGCGGCGATGCTGTGCTCAACGCGGT 4252312

RESULT 11
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 60.8%; Score 61.4; DB 4; Length 4411529;
Best Local Similarity 77.9%; Pred. No. 1.8e-07;
Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 5 GGGCTTGGCCCGAGGGGGGTGCAAGACGTTCTGCGCCGAGCGGAGGCTACGCGG 64
Db 4260150 GTGCTGGCGCCCGAGCGCGGATCAAGTCTTCTACGCGAGCGGCTACACCGC 4260091
QY 65 GCGGAGGGGTGCGCGGTGTGCTCTCAAGCGGCT 99
Db 4260090 TCCGAGGCGCGCGGATGCTGTGCTCAAGCGGCT 4260056

RESULT 12
US-09-060-756-318/C
; Sequence 318, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 318
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-318

Query Match 60.2%; Score 60.8; DB 4; Length 333;
Best Local Similarity 77.1%; Pred. No. 4e-07;
Matches 74; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 5 GGGCTTGGCCCGAGGGGGGTGCAAGACGTTCTGCGCGGAGCGGAGGCTACGCGG 64
Db 182 GCGCTGTACCGACCGCGCGGATGCAAGCTTCTGAGCGGCAACGCGGATGTGCGC 123
QY 65 GCGGAGGGGTGCGCGGTGTGCTCTCAAGCGGCT 100
Db 122 GCGGAGGGGTGCGCGGTGTGCTCTCAAGCGGCT 87

RESULT 13
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/ABPD-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435

```

LOCATION: 8262..9305
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9906..10454
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10707..10964
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 2"
US-07-642-734C-1

Query Match 59.4%; Score 60; DB 1; Length 11219;
Best Local Similarity 75.0%; Pred. No. 5.2e-07;
Matches 75; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Oy 1 GCGCGCGCTTTCGCCGACGGGCGGTGCACAGCTTTCGGCCGACGGGCGCTACGC 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7313 GGGCGCGCTCTTCGCCGACGGCGCGGTGCACGCCCTTTCGGGACGGCGATTGCG 7372
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 61 GCGGGCGAGGGGCGTGGCGCGCTGTGCTGCTCAGGGCGTG 100
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7373 TCTCGGGAGGGGTTCGGCGTTGTGCTGCTCCAGCGTTG 7412
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-08-439-009A-1
Sequence 1, Application US/08439009A
Patent No. 6004787
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
TITLE OF INVENTION: Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven F. Weinstock
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439, 009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952-US-D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea

```

STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc-feature
LOCATION: 744..6659
OTHER INFORMATION: /function="APPROXIMATE SPAN OF
OTHER INFORMATION: MODULE 1"
OTHER INFORMATION: /label= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function="gene="eryA"
OTHER INFORMATION: /product="ORF encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 744..1868
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1998..2198
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 2250..3626
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 3831..4811
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 5574..6125
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 1"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 6369..6626
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain 2 of module 1"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 6678..11219
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 2"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 6678..8066
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 8262..9305
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 2"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 9906..10454
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 10707..10964
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 2"
US-08-439-009A-1
Query Match 59.4%; Score 60; DB 3; Length 11219;
Best Local Similarity 75.0%; Pred. No. 5, 2e-07;
Matches 75; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Result 15
US-08-764-233A-1
Sequence 1, Application US/08764233A
Patent No. 5716849
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Schupp, Thomas
APPLICANT: Beck, James J.
APPLICANT: Hill, Dwight S.
APPLICANT: Neff, Snezana
APPLICANT: Ryals, John A.
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8589
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1, pUJ3, and pVKM15
FEATURE:
NAME/KEY: misc-feature
LOCATION: 383..760
OTHER INFORMATION: /product="SOR"
OTHER INFORMATION: /note="This gene encodes a protein that is highly homolog
OTHER INFORMATION: the reductase domains of type I PKSS such as eryA from
OTHER INFORMATION: Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc-feature
LOCATION: 927..19874

```
OTHER INFORMATION: /product= "SorA"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs that
OTHER INFORMATION: are known to be involved in the synthesis of polypeptide
OTHER INFORMATION: compounds."
FEATURE:
NAME/KEY: misc-feature
LOCATION: 942..715
OTHER INFORMATION: /product= "Module 1 of SorA"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of SorA"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of SorA"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "SorB"
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs genes"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of SorB"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of SorB"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of SorB"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of SorB"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of SorB"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "SorM"
OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
OTHER INFORMATION: polypeptide rapamycin."
US-08-764-233A-1
```

```
Query Match 59.0%; Score 59.6; DB 1; Length 49377;
Best Local Similarity 75.5%; Pred. No. 6.1e-07;
Matches 74; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 GCGGCTTCCCGGAGCGGGGCGGCGGAGAGCTTCGCGGAGCGGAGCGGCTACGGCG 62
Db 1900 GCGGCTTCCCGGAGCGGGGCGGCGGAGAGCTTCGCGGAGCGGAGCGGCTATGTC 1959

QY 63 GCGGAGGGGTGCGCGCTGTGTGTGCTCAAGCGCTG 100
Db 1960 GCGGAGGGGGGCGGCTGTGTGTGCTCAAGCGCTG 1997
```

Search completed: November 5, 2002, 15:39:42
Job time : 1663.74 secs

APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match
Best Local Similarity 47.3%; Score 47.8; DB 10; Length 13842;
Best Local Similarity 67.7%; Pred. No. 8.8e-05;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 1 GCGCGGCTTTGCGCCGAGCGGGGTGCAAGAGCTTCGCGCGAGCGAGCGGTACGC 60
DB 8454 GCGGGGCTCGCGCCGCGAGCGCGCGAGAGGCTTCGCGCGAGCGAGCGGTTCGG 8513
OY 61 GCGGGGCGAGGGGTGCGCGGTGTGTGTCGCAAGCGGCT 99
DB 8514 CCGCGGCGAGGGGCTCGCGCTGCTGCTGAGAGCGGCT 8552

RESULT 3
US-09-861-289-5
Sequence 5, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 36778
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match
Best Local Similarity 47.3%; Score 47.8; DB 10; Length 36778;
Best Local Similarity 67.7%; Pred. No. 8.3e-05;
Matches 67; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 1 GCGCGGCTTTGCGCCGAGCGGGGTGCAAGAGCTTCGCGCGAGCGAGCGGTACGC 60
DB 10195 GCGGGGCTCGCGCCGCGAGCGCGCGAGAGGCTTCGCGCGAGCGAGCGGTTCGG 10254
OY 61 GCGGGGCGAGGGGTGCGCGGTGTGTGTCGCAAGCGGCT 99
DB 10255 CCGCGGCGAGGGGCTCGCGCTGCTGCTGAGAGCGGCT 10293

RESULT 4
US-09-861-289-32
Sequence 32, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.

APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 32
LENGTH: 11220
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match
Best Local Similarity 45.1%; Score 45.6; DB 10; Length 11220;
Best Local Similarity 66.0%; Pred. No. 0.00033;
Matches 66; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 1 GCGCGGCTTTGCGCCGAGCGGGGTGCAAGAGCTTCGCGCGAGCGAGCGGTACGC 60
DB 5262 GCGGGGCTCGCGCGAGGAGCGCGCGGTGCAAGGCTTCGCGCGAGCGAGCGGTTCGG 5321
OY 61 GCGGGGCGAGGGGTGCGCGGTGTGTGTCGCAAGCGGCTG 100
DB 5322 CCGCGGCGAGGGGCTCGCGATGCTGCTGCTGAGAGCGGCTG 5361

RESULT 5
US-09-861-289-36
Sequence 36, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 36
LENGTH: 4041
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-36

Query Match
Best Local Similarity 44.2%; Score 44.6; DB 10; Length 4041;
Best Local Similarity 65.7%; Pred. No. 0.00063;
Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 1 GCGCGGCTTTGCGCCGAGCGGGGTGCAAGAGCTTCGCGCGAGCGAGCGGTACGC 60
DB 747 GAGGGCATGCGCGCGAGCGCGCGAGAGGCTTCGCGCGAGCGAGCGGTTCGG 806
OY 61 GCGGGGCGAGGGGTGCGCGGTGTGTGTCGCAAGCGGCT 99
DB 807 CTGCGGCGAGGGGCTCGCGCTGCTGCTGAGAGCGGCT 845

RESULT 6
US-09-861-289-34
Sequence 34, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.

APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438U1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 4689
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-34

Query Match 43.6%; Score 44; DB 10; Length 4689;
Best Local Similarity 65.0%; Pred. No. 0.0009;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 1 GCGGCGCTTTCGCCGACGCGGCGGTGCAAGAGCTTCGCCGCGACGCGGCGCTACGC 60
DB 753 GCGGCGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 812
OY 61 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 100
DB 813 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 852

RESULT 7
US-09-924-256A-53
Sequence 53, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924,256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 53
LENGTH: 643
TYPE: DNA
ORGANISM: Peltiligera neopolydactyla
US-09-924-256A-53

Query Match 42.4%; Score 42.8; DB 10; Length 643;
Best Local Similarity 68.6%; Pred. No. 0.002;
Matches 59; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 15 CCGACGCGGCGGTGCAAGAGCTTCGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 74
DB 384 CTGATGCTGCTGCAAAACCTTGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 443
OY 75 GCGCGGTGCTGCTGCTCAAGCGGCGCTG 100
DB 444 GTGCTGTTGTAATTCACAAGCGTTTG 469

RESULT 8
US-09-924-256A-69
Sequence 69, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap

APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924,256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 69
LENGTH: 658
TYPE: DNA
ORGANISM: Usnea florida
US-09-924-256A-69

Query Match 41.4%; Score 41.8; DB 10; Length 658;
Best Local Similarity 65.6%; Pred. No. 0.0037;
Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 7 GCTTTCGCCGACGCGGCGGTGCAAGAGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 66
DB 376 GCTTTCCTCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 435
OY 67 CGAGGCGTGCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 99
DB 436 GGAGGCGGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 468

RESULT 9
US-09-924-256A-57
Sequence 57, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924,256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 57
LENGTH: 765
TYPE: DNA
ORGANISM: Peltiligera neopolydactyla
US-09-924-256A-57

Query Match 34.3%; Score 34.6; DB 10; Length 765;
Best Local Similarity 59.8%; Pred. No. 0.027;
Matches 58; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 3 GCGGCGTTCGCCGACGCGGCGGTGCAAGAGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 62
DB 494 GAGGCGCTCTCTCGATGATGCGGTGCTACCTTCGATGATGATGATGATGATGATGATGAT 553
OY 63 GGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 99
DB 554 GGGGAGAGGTGTCGCGCACATCGTTTCAACGCGCT 590

RESULT 10
US-09-924-256A-79
Sequence 79, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian

```

; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; PRIOR FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 79
; LENGTH: 761
; TYPE: DNA
; ORGANISM: Uanea florida
US-09-924-256A-79
```

```
Query Match
Best Local Similarity 34.1%; Score 34.4; DB 10; Length 761;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
QY 8 CTTTGGCCCGACGGCGGTGCAAGAGTTCGGCCGACGGGAGCGCTACGGCGGGC 67
DB 495 CTTTGGCCCGACGGCGGTGCTTGTGATCAACGCTGTAATGTTATGCGCGTGG 554
QY 68 GAGGGGTGGCGGTGCTGCTCAAGCGCT 99
DB 555 GAGGAGACGGCGGTGCTGCTCAAGCGCT 586
```

```
RESULT 11
US-09-960-352-175
; Sequence 175 Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tso, Mengling
; APPLICANT: Bysit, John C.
; APPLICANT: Mathias, Nagappan
; TITLE OF INVENTION: MUSCLE ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511006.32(1029)C
; CURRENT APPLICATION NUMBER: US/09/960 352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 175
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (386)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 01-LIB34-045-01-EL-A1
US-09-960-352-175
```

```
Query Match
Best Local Similarity 32.3%; Score 32.6; DB 10; Length 435;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
```

```
QY 23 CGGTGCAAGAGCTTCGGCGCGACGCGCTACGCGCGGGCGAGGGTGGCCCTG 82
DB 186 CMTGCGCGCGGTTCGGCGCGACGCTGCGCTACGCGCGGGAGAGCGCGGGCTG 245
QY 83 GGTGCTCAAGCGGGCTG 101
DB 246 GCGGCGGTGCAACATCTG 264
```

```
RESULT 12
US-09-924-256A-29
; Sequence 29 Application US/09924256A
; Patent No. US20020127659A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; PRIOR FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 29
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Clone
US-09-924-256A-29
```

```
Query Match
Best Local Similarity 32.3%; Score 32.6; DB 10; Length 643;
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```
QY 14 CCGAGCGGCGGTGCAAGAGCTTCGCGCGACCGGAGCGCTACGCGGGCGGAGGG 73
DB 363 CCGAGCGGCGGTGCAAGAGCTTCGCTGATGAGCTGCTGATGAGCTGAGGAGGA 442
QY 74 TCGCGCGGTGCTGCTCAAGCGGCTG 100
DB 443 GTTGTGCTGCTGCTCAAGCGCTTGG 469
```

```
RESULT 13
US-09-820-893-37
; Sequence 37 Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: P2033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; CURRENT FILING DATE: 2001-03-30
; PRIOR FILING DATE: 2000-03-20
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1466
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-820-893-37
```

```
Query Match
Best Local Similarity 31.9%; Score 32.2; DB 10; Length 1466;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
```

```
QY 16 CGAGCGGCGGTGCAAGAGCTTCGCGCGACCGGAGCGCTACGCGGGCGGAGGGTG 75
DB 443 CGCGCGCTCTTCAGAGATACACAGTGGAGCACCGCTACTGGGCGGAGCGGCTG 502
QY 76 CGCGGTGCTGCTCAAGAGCTG 100
DB 503 GAGCTGAGAGTGGCCACGCGCTG 527
```

```
RESULT 14
US-09-764-869-2349
```

```
; Sequence 2349, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2349
; LENGTH: 13273
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-2349
```

```
Query Match 31.9%; Score 32.2; DB 10; Length 13273;
Best Local Similarity 57.4%; Pred. No. 0.95;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```

```
OY 1 GCGGCGCCTTTCGCCGACGGCGGTGCAAGACGTTTCGCCGCCGACGGCGCTACGC 60
Db 1884 GCGCCGCGCTCCGCTCCAGGTGCTGTGACGCGCGCTGCTACGC 1943
```

```
OY 61 GCGGGGCGAGGGGTGCGCGTGTGCTGCTCAAGCGGCTGC 101
```

```
Db 1944 GCTGCTGGCGCTGAAGAGAGTGAGAGATCATGTTGCTGC 1984
```

RESULT 15

```
US-09-924-256A-49
; Sequence 49, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 722
; TYPE: DNA
; ORGANISM: Parmelia sulcata
US-09-924-256A-49
```

```
Query Match 31.1%; Score 31.4; DB 10; Length 722;
Best Local Similarity 57.5%; Pred. No. 1.8;
Matches 50; Conservative 3; Mismatches 34; Indels 0; Gaps 0;
```

```
OY 7 GCTTTCGCCGACGGCGGTGCAAGACGTTTCGCCGCCGACGGCGCTACGCGTGGAG 66
Db 443 GTTGTCGCCCACTGCGCGATGTGCGAGCGCTAACSGGATGCTATGCATGAGG 502
```

```
OY 67 CGAGGGGTGCGCGCTGTGTGCTCAA 93
```

```
Db 503 AGAGGGTATGTGATCTGTGCTCAA 529
```

Search completed: November 5, 2002, 23:01:17
Job time : 97.2537 secs

11

—

—


```
source
1. .516
/organism="Drosophila melanogaster"
/db_xref="taxon:7127"
/clone="SD03777"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pot2"
/lab_host="DH5-alpha"
/notes="Vector: pMT2; Site1: EcoRI; Site2: XhoI; Sized
fractionated cDNA were directly ligated into pot2.
Plasmid cDNA library."
BASE COUNT      101 a      163 c      158 g      94 t
ORIGIN
Query Match      41.4%; Score 41.8; DB 9; Length 516;
Best Local Similarity 68.2%; Pred. No. 2;
Matches 58; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 14 CCCGACGGCGGTCGACAGCTTCTCGGCAGACGGAGCGCTACGCGCGGCGAGGGG 73
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 201 CACGACGGCGCTCTGCAAGACCTTCGATGCACGCGCAAGCGCTATGCCGCGCGAGCACC 260
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 74 TGGCGCGTGGTGGCTCAAGCGGC 98
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 TGGCGCTGTCTGCTCTCAGAGGC 285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
BM488815      59' bp      mRNA      linear      EST 07-FEB-2002
LOCUS      BM488815
DEFINITION      Epiphyseal Growth Plate cDN Library (pgm2n) Gallus gallus cDNA
      clone pgm2n.pk008.p20 5' similar to ref1p_46119.1 (NM_053667)
      growth suppressor 1 leprecan [Rattus norvegicus]
      db|AD51875.1|AF087433.1 (AF087433) leprecan [Rattus norvegicus],
      mRNA sequence.
ACCESSION      BM488815
VERSION      BM488815
KEYWORDS      EST.
SOURCE      Chicken
      Gallus gallus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
      Phasianinae; Gallus.
      1 (bases 1 to 597)
REFERENCE      Cogburn,L.A. and Monsonego-O'Ryan,E.
      ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and
      Epiphyseal Growth Plate cDNA Library, USDA/IRAFs Animal Genome
      Project
      Unpublished (2002)
JOURNAL      Contact: Larry A. Cogburn
      University of Delaware
      Townsend Hall, Newark, DE 19717, USA
      Tel: 302-831-1335
      Fax: 302-831-2822
      Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES
source
1. 597
/organism="Gallus gallus"
/strain="Commercial broiler and Ottawa Res. Centre
      Strains 90 & 21"
/db_xref="taxon:9031"
/clone="pgm2n.pk008.p20"
/clone_lib="Normalized Chicken Breast Muscle, Leg Muscle,
      and Epiphyseal Growth Plate cDNA Library (pgm2n)"
/sex="Male and Female"
/tissue_type="Breast muscle, leg muscle and epiphyseal
      growth plate"
/dev_stage="Breast,leg:Embryo(d19);post-hatch(1d,1,3,5,7,9
      ,11 weeks);growth plate(1d,7d,14d post-hatch)"
/lab_host="E. coli DH10B"
/notes="Vector: PCWMS10R6; Library made from equivalent
      pools of total RNA isolated from each tissue (embryonic
      muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
```

```
plate 33.3% of the final RNA pool). Single pass sequencing
      from 5' end"
BASE COUNT      78 a      200 c      199 g      90 t      30 others
ORIGIN
Query Match      40.4%; Score 40.8; DB 13; Length 597;
Best Local Similarity 63.0%; Pred. No. 3.4;
Matches 63; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 1 GCGCGCGCTTTCGCCCGACGGCGGTCGCAAGACCTTCTGCGCGAGCGGCTACGC 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68 GCGCGCGCTGCGCGCGAGCGCCCGGAGCGCGCTCTGCTGCGCGCGCGAGCGGCGG 127
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GCGGCGGAGGGTGGCGCGCGTGGTGGCTCAAGCGCGTG 100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 128 GCGGGGGAGCTGGCGCGCGCTGCTGCAAGTGAAGCGG 167
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
BH179086      738 bp      DNA      linear      GSS 19-OCT-2001
LOCUS      BH179086/c
DEFINITION      013.D.19-rev SMBAC1 Schistosoma mansoni genomic clone 013D19 5',
      DNA sequence.
ACCESSION      BH179086
VERSION      BH179086.1 GI:16279565
KEYWORDS      GSS.
      Schistosoma mansoni.
SOURCE      Schistosoma mansoni.
      Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
      Strigeiida; Schistosomatoidea; Schistosomatidae; Schistosoma.
      1 (bases 1 to 738)
REFERENCE      Le Paslier,M.C., Pierce,R.J., Mellin,F., Hiral,H., Wu,W., Williams
      ,D.L., Johnston,D., Loye-de,F.T. and Le Paslier,D.
      Construction and characterization of a Schistosoma mansoni
      bacterial artificial chromosome library
      Genomics 65 (2), 87-94 (2000)
      20247247
      Other_GSSs: 013.D.19-21
JOURNAL      INSERM U 167
      Contact: Pierre RJ
MEDLINE      Institut Pasteur de Lille
      1 rue du Professeur A. Calmette, 59019-Lille, France
      Tel: (33) (0)3 20877783
      Fax: (33) (0)3 20877888
      Email: Raymond.Pierce@pasteur-lille.fr
      CNS sequencing ID=DC0A013CB10BP1
      Plate: 013 row: D column: 19
      Seq primer: M13 reverse primer
      Class: BAC ends
COMMENT      High quality sequence stop: 738.
FEATURES
source
1. 738
/organism="Schistosoma mansoni"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="013D19"
/clone_lib="SMBAC1"
/sex="mixed"
/dev_stage="cercariae"
/lab_host="Biomphalaria glabrata"
/notes="Vector: pBelOBAC 11; Site1: Hind III; Partially
      Hind III digested and size-selected S. mansoni cercarial
      DNA was ligated into Hind III digested pBelOBAC 11 vector
      and used to transform E. coli DH10B. The complete library
      contains 23808 clones from 4 independent
      sizing-ligation-transformations. Average insert size
      ranges from 70-127 kb and genome coverage is 7.9-fold."
BASE COUNT      239 a      147 c      124 g      168 t
ORIGIN
Query Match      35.8%; Score 36.2; DB 17; Length 738;
Best Local Similarity 32.7%; Pred. No. 37;
Matches 33; Conservative 38; Mismatches 30; Indels 0; Gaps 0;
```

OY	1	GGCGGCGGTTCGCCCCAGCGGGCGGGTGCAGAAGCGTTCTCAGCGACGCGGACGGCTATGCG	60
Dd	118	SSGSGSGSTCCGSGSGSSSSSSSGSGGGGTTCGGGSGGSGGSSSSSGSGSGGSGGSGG	59
OY	61	GGCGGCGAGGAGGAGTGCGCCGTGCTGCTCAAGCGGCTGC	101
Dd	58	GGGSGSGSGSGGGGSSSGSKGKKKGKTGTTCCGSBBBGB	18
RESULT 4			
CNSO7L94/c		738 bp	DNA
LOCUS	T3	end of clone O13CB10 of library SmbAC1 from strain Puerto-Rican	linear GSS 08-OCT-2001
DEFINITION	T3	end of clone O13CB10 of library SmbAC1 from strain Puerto-Rican	
ACCESSION	AL616042		
VERSION	AL616042.1	GI:16029266	
KEYWORDS	GSS.		
SOURCE	Schistosoma mansoni.		
ORGANISM	Schistosoma mansoni		
REFERENCE	Eukaryote; Metazoa; Platyhelminthes; Trematoda; Digenea;		
AUTHORS	Strigalidda; Schistosomatidae; Schistosomatidae; Schistosoma.		
TITLE	1 (bases 1 to 738)		
JOURNAL	Le Paslier,M.C., Pierce,R.J., Merlín,F., Hirai,H., Wu,W., Williams,D.L., Johnston,D., Loverde,P.T. and Le Paslier,D.		
MEDLINE	Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library		
PUBMED	Genomics 65 (2), 87-94 (2000)		
REFERENCE	20247247		
AUTHORS	10783255		
TITLE	2 (bases 1 to 738)		
JOURNAL	Genoscope.		
COMMENT	Direct Submission. Submitted (05-OCT-2001) Genoscope - Centre National de Séquencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr Partially Hind III digested and size-selected 5' .mansoni cercarial DNA was ligated into Hind III digested pBlotBAC 11 vector and used to transform E. coli DH10B. The complete library contains 23808 clones from 4 independent sizing-ligation-transformations. Average insert size ranges from 70-127 kb and genome coverage is 7.9-fold.		
FEATURES	Location/Qualifiers		
SOURCE	1..738		
BASE COUNT	239 a 147 c 124 g 168 t 60 others		
ORIGIN	/organism="Schistosoma mansoni" /strain="Puerto-Rican" /db_xref="taxon:6183" /clone="013CB10" /clone_id="SmbAC1" /note="end : T3"		
Query Match	35.8%, Score 36.2, DB 17; Length 736;		
Best Local Similarity	32.7%, Pred. NO. 37;		
Matches	33; Conservative 38; Mismatches 30; Indels 0; Gaps 0;		
OY	1	GGCGGCGGTTCGCCCCAGCGGGCGGGTGCAGAAGCGTTCTCAGCGACGCGGACTAGCG	60
Dd	118	SSGSGSGSTCCGSGSGSSSSSGSGGGGTTCGGGSGGSGGSSSSSGSGSGGSGGSGG	59
OY	61	GGCGGCGAGGAGGAGTGCGCCGTGCTGCTCAAGCGGCTGC	101
Dd	58	GGGSGSGSGSGGGGSSSGSKGKKKGKTGTTCCGSBBBGB	18
RESULT 5			
LOCUS	Bf182804	898 bp	mRNA
DEFINITION	601809319F1 NIH_MGC_18 Homo sapiens CDNA clone IMAGE:4040045 5', mRNA sequence.		linear EST 31-OCT-2000
ACCESSION	Bf182804		
VERSION	Bf182804.1	GI:11060947	
KEYWORDS	EST.		

[illegible]

OY		1	GCGGCGCTTTTCGCCACGCAGCCGGTGTGAACGTTTTCCGCGCAGCGAAGCGGTATACG	60
Dd		785	GCGCGCGCCCGCCGCGCGTGCGCGCGGGGGCGGCGCCTCTCGGCGCGGTCCGCGCGCCC	844
OY		61	GCGGCGCGAGGGGCTGC	76
Dd		845	GCGGCGCGCGCGTGC	860
RESULT 9	BH611949			
LOCUS	BH611949		463 bp	DNA linear GSS 04-JAN-2002
DEFINITION	SALK_031928 Arabidopsis thaliana TDNA insertion lines Arabidopsis thailiana genomic clone SALK_031928, DNA sequence.			
ACCESSION	BH611949			
VERSION	BH611949.1		GI:18059245	
KEYWORDS	GSS.			
SOURCE	thalae cross. Arabidopsis thaliana			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi Alonso,J.M., Leisner,T.J., Bataejas,P., Chen,H., Cheuk,R., Gadrihab, , C., Teske,A., Karnes,M., Kim,C.U., Parker,H., Prednits,I., Shinn,P., , Zimmerman,J., and Ecker,J.R. A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome Unpublished (2001) Contact: Joseph R. Ecker Salt Institute Genomic Analysis Laboratory (Signal) The Salt Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel.: 858 453 4100 x1752 Fax: 858 558 6379 Email: jekeres@salt.edu This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged. Location/Qualifiers <pre> 1..463 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone_id="SALK_031928" /note="PCR was performed on Arabidopsis thailiana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/Tdna_protocols.html"</pre>			
BASE COUNT	63 a	111 c	190 g	71 t
ORIGIN				28 others
Query Match		33.9%	Score 34.2;	DB 17; Length 463;
Best Local Similarity	59.4%;	Pred. No. 1.1e+02;	Matches 57;	Conservative % 0; Mismatch 39; Indels 0; Gaps 0;
OY		3	GCGCGCTTTTCGCCACGCAGCGGTCTCGGCGGAGCGGACTTAAGCGGC	62
Dd		351	GCGCGCGCCCGCCCGCGGGGGGGGGGGGGGGGGGGGGCGGCGCGCGGTGGCGGCC	410
OY		63	GCGGCGAGGGTGC GCCGTGTGAGTCAAGCGGC	98
Dd		411	GCGGCGGGGGGGCGCGCGCGCGNCGCGCGGC	446
RESULT 10	BFA29029		470 bp	mRNA linear EST 29-NOV-2000
LOCUS	BFA29029			
DEFINITION	WHEI1704_C03.F06GS wheat heat stressed spike cdna library Triticum aestivum CDNA clone WHEI1704_C03.F06, mRNA sequence.			

FEATURES	SOURCE
REFERENCE	BF429029
AUTHORS	Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Malatrasi, M., Miller, R., Nguyen, H.T., Rausch, C.J., Seaton, C.L., Tong, J.C., and Zhang, D.
JOURNAL	US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
COMMENT	800 Buchanan Street, Albany, CA 94710, USA
TITLE	Heat stressed spike cDNA library
UNPUBLISHED	(2000)
GENOMES	Heat stressed spike cDNA library
SPERMATOPHYTES	Tracheophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae
TRITICEAE	Triticum
TRITICUM	(bases 1 to 470)
TRITICUM	Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Malatrasi, M., Miller, R., Nguyen, H.T., Rausch, C.J., Seaton, C.L., Tong, J.C., and Zhang, D.
TRITICUM	The structure and function of the expressed portion of the wheat
TRITICUM	genomes - Heat stressed spike cDNA library
TRITICUM	Unpublished (2000)
TRITICUM	Contact: Olin Anderson
TRITICUM	US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
TRITICUM	800 Buchanan Street, Albany, CA 94710, USA
TRITICUM	Tel: 510595773
TRITICUM	Fax: 510595818
TRITICUM	Email: oanderson@wp.usda.gov
TRITICUM	Sequence have been trimmed to remove vector sequence and low
TRITICUM	quality sequence with phred score less than 20
TRITICUM	Seq primer: Stragene SK primer.
TRITICUM	Location/Qualifiers
TRITICUM	1..470
TRITICUM	/organism="Triticum aestivum"
TRITICUM	/cultivar="Chinese Spring"
TRITICUM	/db_xref="taxon:4565"
TRITICUM	/clone="WHE1704.C03.F06"
TRITICUM	/clone_id="Wheat heat stressed spike cDNA library"
TRITICUM	/lisse_type="Whole spike"
TRITICUM	/dev_stage="Spikes at 5, 10, 15 and 20 days after
TRITICUM	anthesis"
TRITICUM	/lab_host="E. coli SOLR"
TRITICUM	/note="Vector: Lambda Uni-ZAP XR, excised phagemid.
TRITICUM	Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20
TRITICUM	days after anthesis were heat stressed under two
TRITICUM	conditions at Texas Tech University (D. Zhang in HT Nguyen
TRITICUM	lab): (1) at 38 C for 4 hours and (2) 5 days of cyclic
TRITICUM	treatment of 38 C for 4 hours. Total RNA and poly(A) RNA
TRITICUM	were prepared, a cDNA library was made, and the cDNA
TRITICUM	clones were in vivo excised to give plasmid phagemids
TRITICUM	in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at
TRITICUM	the University of California, Riverside. Plasmid DNA
TRITICUM	preparations and DNA sequencing were performed in the OD
TRITICUM	Anderson lab (all other authors)."
TRITICUM	BASE COUNT
TRITICUM	46 a 220 c 138 g 65 t 1 others
TRITICUM	ORIGIN
TRITICUM	Query Match
TRITICUM	33.7%; Score 34; DB 12; Length 470;
TRITICUM	Best Local Similarity 58.6%; Pred. No. 1.2e+02;
TRITICUM	Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
TRITICUM	DB
TRITICUM	2 CGCGCGCTTTGCGCCGAGAGGGGCGGTCTCGCGCGAGCGAGCGGTACGCG 61
TRITICUM	
TRITICUM	DB 350 CACGCGCGTGGCGGTGCGGGGGGAGCAGNGTGGTGGGAGGAGCGCTCGCGCG 409
TRITICUM	
TRITICUM	DB 62 CGGGCGAGAGGGGTGCGCGGTGTGCTCAAGCGGCTG 100
TRITICUM	
TRITICUM	DB 410 GCGAGGAGCGCTCGCGCGAGCTGTGGAGTGGCGCCCTG 448
TRITICUM	
TRITICUM	RESULT 11
TRITICUM	AL564234/c
TRITICUM	LOCUS
TRITICUM	AL564234 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSODM003YF10 3
TRITICUM	DEFINITION
TRITICUM	prime, mRNA sequence.
TRITICUM	ACCESSION
TRITICUM	AL564234
TRITICUM	VERSION
TRITICUM	AL564234.1 GI:12914435
TRITICUM	EST.
TRITICUM	human

ORGANISM	<i>Homo sapiens</i>
EXTRACT	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 559)
AUTHORS	Li, W. B., Gruber, C., Jesse, J., and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope

BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
location/Qualifiers:

FEATURES
source

```

//issue_type="neuroblastoma cells"
/lab_host="DHI08"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a Kc1t-Oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact: Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax: (1) 301 610
8371 Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com"
69 a 327 c 97 g 44 t 22 others

```

Query Match	33.7%	Score 34	DB 9	length 559
Best Local Similarity	57.4%	Pred. No	1.2e+02	
Matches	58	Conservative	1	Mismatches 42; Indels 0; Gaps 0;

```

QY      1  GCGCGCGCTTTCGCGCCGACGGCGGTGCAAI;ACGTTCTCGCGCCGACGGCGGCTACGC 60
      1  ||||  |||  : ||||  ||  ||  ||  ||  ||  ||  ||
Db      440  GGNCGCGCGCGCGCGSGGCGGGGCGCGG;GCGGGGTGGGGGGGCGCGCTGCGGGGCGG 381

```

Oy	61 GCGGGGAGGGGTGCCCTGTGTCTTAAAGGGCTGC	101
Dd	380 GCCGGGGGGGGGGGGGGGGCCGAAGGGGGCGGC	340

RESULT 12				
AZ185516/c				
LOCUS	AZ185516	775 bp	DNA	linear
DEFINITION	SP.1005, A1-H12, SP66 Strongylid-centroctus purpuratus, purple sea urchin, sperm genomic BAC library Strongylid-centroctus purpuratus genomic clone Plate=1005 COL=23 Row=O, DNA sequence.			
ACCESSION	AZ185516			

REVISION 42105210
VERSION A2185516.1 GI:8357801
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus

REFERENCE

TITLE	A sea urchin genome project: sequence scan, virtual map, and additional resources
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
COMMENT	20402566
	Contact: Cameron, RA, Davidson, EH, Hood, L

California Institute of Technology
Pasadena California 91125, USA.

FEATURES
SOURCE
 Tel: (626) 395-8421
 Fax: (626) 793-3047
 Email: acameron@caltech.edu
 Plate: 1005 row: 0 Column: 23
 Seq primer: SP6
 Class: PAC ends
 High quality sequence stop: 775.
 Location/Qualifiers
 1..775

BASE COUNT ORIGIN	46 a	451 c	243 g	27 t	8 others
----------------------	------	-------	-------	------	----------

Query Match	33.7%	Score 34	DB 17	Length 775
Best Local Similarity	66.2%	Pred. No.	1.2e+02	
Matches 49	Conservative 0	Mismatches 25	Indels 0	Gaps 0

Oy 13 GCCGACGGCGGCTGCAGACGTTCCTCGGCCGACGCGGACTACGGCGGGGGCAGAAGG 72
| | | | | | | | | | | | | | | | | |
Db 303 GCGCGCCGGCGCGCGCGCGCGGCGTCTCGGCGCGCGCGCGCGCGCGCGGGGGGGGCGCGGG 214

QY	73	GTGCGCGTGTGG	86
Db	243	GCGGCGGCGGGG	230

RESULT	13
BM910915/c	
LOCUS	
DEFINITION	BW10915 1096 bp mRNA linear EST 12-MAR-2002
ACCESSION	AGNCOURT_6615874 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454494
DESCRIPTION	5' mRNA sequence.

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 1096)				
	NIH-MGC http://mgc.nci.nih.gov/				
	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				

CDNA Library Arrived by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LLCM1949 row: 1 column: 15
High quality sequence stop: 590.
Location/Qualifiers

FEATURES
SOURCE

```

"organism":"Homo sapiens"
/db_xref="taxon:9606"
"clone_image":5454494"
"clone.lib":"NH1.MCC.98"
"tissue_type":"astrocytoma grade IV, cell line"
"lab_host":"DH10 (phage-resistant)"
"note":"CDNA made by brain; Vector: pCR7; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adapter: GGCACCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald N. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit

```

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 221 a 380 c 304 g 190 t 1 others

Query Match 33.7%; Score 34; DB 14; Length 1096;
Best Local Similarity 66.2%; Pred. No. 1.1e+02;
Matches 49; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Y 2 CGCGCGCTTTGGCCGAGCGGCTGCAAGACGTTCTGCGCCGAGCGGCGCTACGCG 61
DB 628 CGCTTGCTCTGCTGCTGCGAGGTGAAGAAGTCCCGACGCTTCAACTGCGAGGCC 569
Y 62 CGGCGCGAGGCGTG 75
DB 568 AGGGCGGAGGCGTG 555

RESULT 14
B0468858/c 155 bp mRNA linear EST 30-MAY-2002
LOCUS HM02J057 HM Hordeum vulgare cDNA clone HM02J05 5-PRIME, mRNA
DEFINITION
ACCESSION B0468858
VERSION B0468858.1 GI:21276640
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 155)
Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
EST sequencing and analysis in barley (2002)
Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 155 Std Error: 0.00
Plate: 2 row: J column: 5
Seq primer: M13rev.

REFERENCE
AUTHORS Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
TITLE EST sequencing and analysis in barley (2002)
JOURNML Unpublished (2002)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 155 Std Error: 0.00
Plate: 2 row: J column: 5
Seq primer: M13rev.

FEATURES
Source
Location/Qualifiers
1..155
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HM02J05"
/clone_1lb="HM"
/tissue_type="male inflorescences"
/dev_stage="male inflorescences (approx. 2 mm in size),
green anther stage"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+, site 1: EcoRI (5'-end of
cDNA); site 2: XhoI (3'-end of cDNA); Due to a cloning
artefact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable."

BASE COUNT 14 a 62 c 65 g 14 t

Query Match 33.5%; Score 33.8; DB 14; Length 155;
Best Local Similarity 62.4%; Pred. No. 1.5e+02;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
Y 14 CCGGACGGCGGTGCAACGTTCTGCGCCGAGCGGCTACGCGCGGCGGAGGG 73
||||| | ||| | |||| ||||| || ||||| |||||

DB 101 CCGGACGGCTGTGCGGGGTGTGCAACCGGAGCGGAGCGGCCCGCGGACGCTC 42
Y 74 TCGCGCGTGTGTGCTGCTCAAGCGC 98
DB 41 TCGCGCGTGTGCGCGCGCTACGCGC 17

RESULT 15
LOCUS BE903754 660 bp mRNA linear EST 29-SEP-2000
DEFINITION 601675594P1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958415 5',
mRNA sequence.
ACCESSION BE903754
VERSION BE903754.1 GI:10395293
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.nhl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNML Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rtmail.nih.gov
Tissue Procurement: ARCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at: image.lnl.gov
Plate: LICM837 row: k column: 24
High quality sequence start: 3
High quality sequence stop: 660.
Location/Qualifiers
1..660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3958415"
/clone_1lb="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

FEATURES
Source
Location/Qualifiers
1..660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3958415"
/clone_1lb="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 145 a 163 c 200 g 152 t

Query Match 33.5%; Score 33.8; DB 12; Length 660;
Best Local Similarity 58.4%; Pred. No. 1.3e+02;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Y 1 CGCGCGCTTTGCGCCGAGCGGCGGTGCAAGACGTTCTGCGCCGAGCGGCGGCTACGCG 60
DB 9 GGGGAGAGTGGCGCATGATGCTGTTCACAGCGGTGCTGTTGAGCACGCGGTGCGCTACG 68
Y 61 CGGCGGCGAGGCGGTGCGCGCGTGTGCTGCTCAAGCGGCGCTGC 101
DB 69 GCTGCTGCGCTGAGAGGAGTGAAGATCATGCTGCTGC 109

Search completed: November 5, 2002, 16:18:09
Job time : 1117.44 secs

AUTHORS Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE Cloning and heterologous expression of the epochllone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE Isolation and characterization of the epochllone biosynthetic gene cluster from *Sorangium cellulosum*
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien,B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Korean Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES
source 1..58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
1..992
/note="orfA"
/codon_start=3
/product="putative transposase"
/protein_id="AAFG2876.1"
/db_xref="GI:7453535"
/translation="VRHTEAFADGVPRLVLDNLKNAVERGDAIRHPITLAL
SADYRFEPRPAVARGNE:GVERAIRVRGFEFARAYADIDNRATETTSAL
DRSWEDARVYRQAFDI:RSVLLRHPTPEPDHERVEQKPTARPAIDVSPH
DTRRTLVADLSQVRI:DGNOIVATVHRSDRGQOIEQPHLORLYDEKRAEHR
GLDRLAARASSQAFLRIYAEKDGWGSALNLLQLDVAAGAALEALVEYLEDTI
HIGAVROYIDRRSEERHLIPVSIPTVGEHNAIVVTEHSITTYDALKKDPTP"
989..1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAFG62879.1"
/db_xref="GI:7453555"
/translation="MTDLTPETTKDLKSLGFLGLACMEQADKPKMLREVLAIERE
RHKSLERLRKNSRYAARF:PMTPDSSMPKRIIDREAVDLDVSRVADLDFEVYTRVD
AOKPLLSTNKAFAWMGOV:PHAACVTVLVRLVRAEVIIEAESYRLKEKELNAT
RTKORRTKKH"
1998..6263
/gene="epoA"
1998..6263
/note="type I polypeptide synthase; epochllone PKS loading module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAFG62880.1"
/db_xref="GI:7453557"

gene
CDS
/translation="MADRPIERAEDPIAIVGACGRPLDGPVLDLSEGFMTLLGSGHDY
GOVPERMDAAMPDDLDGPKTPYTRASF:SDVACDASFPGISPREALMDPAHR
LLEVCWEALNMAALAPALVGTETGVITGIPSEYEAALPAPATASAEIDAHGIGTM
PSVAGRI:SYVLGLRGPCVA:VDYATSSSLVAVHLACOSLRSGECSTALAGVSLMLSP
STLWLSKTRALATDGRCKA:SAEADGFGREGCAVVVLRKLSGRAGADRLIATIRG
SAINHDGSGGLTVNGSSO:ITVLKRALDAGCAASVYLRAGHGTGTLIDPIETOA
LNAYVGGADVATPLLIGSV:CTNIGHPYASGTIGLLKVVLSIOHGOTPAH:HAOLN
PRISGDLRTVTRAPALDAHALVIRHDEITYPSQICGAYPSIATTRMSAHEHRLVAT
RABLLYLSARPAALDAHALVIRHDEITYPSQICGAYPSIATTRMSAHEHRLVAT
SSGCRANDAAAGQTPPGYVRGTDSSRGKALFLTGQGOITLGMKRGIDVWPAF
KEAFELVLEHNOELDRPLRI:VMAHEPASVDALIDOTFTQPALFTEYVLAALMRS
WGEPELVLAGHSIGELVAC:AGVFSLEDAVLVAKRRLQWALPAGGAMVSTIAPEA
DVAAVAPHAASVSTIAAVNG:IDQVITAGAGPVHAI:AAAMA:RGARTALVSHAFHS
PLMA:MLEAFGRVAESVYRI:PSIVLVSNLSGKAGTDEVSFGVWVRHARAEVVRADG
VKALHAAGTFFVEVGRKSTI:LGIVPAQLP:APARALLASSRAGRDEPATVLEALGML

gene
CDS
/translation="MEBOESSAIVIVGMSGFGPARGARDLDEFWRNLRDGTAEVQFSEQ
BLASGVPAVLVLPNRYRAGSVLEVDYRFPAAFGFISPRAEALMDPDRIFEMCAE
ALENAGYPTAYEESIGVYAGANNSSYITSNLHEHPAMRPGMFOQLIGNDKGLAT
HVSYLLNRGSI:SVOTACSTSLVAVHLACOSLRSGECSTALAGVSLMLSP
AEGGIFSPDGCRAPDAKANGITMNGCGVLLLDRECDMALAGIITVRI:PHRAGYV
GARKIFGPAPEVGOAOMALMEALVAGEASIOYIEHGGTGLDGLA:TEAALRAF
DRDSTRSCAIGSVKTSIGHLESAGAGI:IKTVL:LEHQDLP:PSLNFESPN:SIDF
AASPPYVTSIKDMWNGSTPRRAVCSFGSGTGTANAVYLERAPAKALPAA:PARSAL
FVYSKSAALDAAARIRHLOHOGISLIDVAFLATATSPENHRLA:MAAPREAL
REGDAAARQSTPGAVRGRSBCGVYVYVFPFGOSQWVCMKROLIAEPYRHAL
SACDRATIREKMSLILBLADBSGQIERIDVOQVFLVAVPALMRKMGCPDY
VIGHSGVSAVNSRPSITVLSGEPAIGVSLNKRISGOSGEMAVTESHBAALR
GIDLELAAGLGRPRAAIVPMRSTVIGAMVAGELCANWNNLRDPPVRAEYVQAOLOG
EGHLEVEFSPHPIITTSVEEMRRRAAGAGVGSILRRQDSTPRM:LEALGLAMOGP
VPWGRILFPAGGRVPLPTYPWOREYVTEADPAKSNAGDRGRVAGGHLIGEMOTLST

gene
CDS
/translation="MEBOESSAIVIVGMSGFGPARGARDLDEFWRNLRDGTAEVQFSEQ
BLASGVPAVLVLPNRYRAGSVLEVDYRFPAAFGFISPRAEALMDPDRIFEMCAE
ALENAGYPTAYEESIGVYAGANNSSYITSNLHEHPAMRPGMFOQLIGNDKGLAT
HVSYLLNRGSI:SVOTACSTSLVAVHLACOSLRSGECSTALAGVSLMLSP
AEGGIFSPDGCRAPDAKANGITMNGCGVLLLDRECDMALAGIITVRI:PHRAGYV
GARKIFGPAPEVGOAOMALMEALVAGEASIOYIEHGGTGLDGLA:TEAALRAF
DRDSTRSCAIGSVKTSIGHLESAGAGI:IKTVL:LEHQDLP:PSLNFESPN:SIDF
AASPPYVTSIKDMWNGSTPRRAVCSFGSGTGTANAVYLERAPAKALPAA:PARSAL
FVYSKSAALDAAARIRHLOHOGISLIDVAFLATATSPENHRLA:MAAPREAL
REGDAAARQSTPGAVRGRSBCGVYVYVFPFGOSQWVCMKROLIAEPYRHAL
SACDRATIREKMSLILBLADBSGQIERIDVOQVFLVAVPALMRKMGCPDY
VIGHSGVSAVNSRPSITVLSGEPAIGVSLNKRISGOSGEMAVTESHBAALR
GIDLELAAGLGRPRAAIVPMRSTVIGAMVAGELCANWNNLRDPPVRAEYVQAOLOG
EGHLEVEFSPHPIITTSVEEMRRRAAGAGVGSILRRQDSTPRM:LEALGLAMOGP
VPWGRILFPAGGRVPLPTYPWOREYVTEADPAKSNAGDRGRVAGGHLIGEMOTLST

gene
CDS
/translation="MEBOESSAIVIVGMSGFGPARGARDLDEFWRNLRDGTAEVQFSEQ
BLASGVPAVLVLPNRYRAGSVLEVDYRFPAAFGFISPRAEALMDPDRIFEMCAE
ALENAGYPTAYEESIGVYAGANNSSYITSNLHEHPAMRPGMFOQLIGNDKGLAT
HVSYLLNRGSI:SVOTACSTSLVAVHLACOSLRSGECSTALAGVSLMLSP
AEGGIFSPDGCRAPDAKANGITMNGCGVLLLDRECDMALAGIITVRI:PHRAGYV
GARKIFGPAPEVGOAOMALMEALVAGEASIOYIEHGGTGLDGLA:TEAALRAF
DRDSTRSCAIGSVKTSIGHLESAGAGI:IKTVL:LEHQDLP:PSLNFESPN:SIDF
AASPPYVTSIKDMWNGSTPRRAVCSFGSGTGTANAVYLERAPAKALPAA:PARSAL
FVYSKSAALDAAARIRHLOHOGISLIDVAFLATATSPENHRLA:MAAPREAL
REGDAAARQSTPGAVRGRSBCGVYVYVFPFGOSQWVCMKROLIAEPYRHAL
SACDRATIREKMSLILBLADBSGQIERIDVOQVFLVAVPALMRKMGCPDY
VIGHSGVSAVNSRPSITVLSGEPAIGVSLNKRISGOSGEMAVTESHBAALR
GIDLELAAGLGRPRAAIVPMRSTVIGAMVAGELCANWNNLRDPPVRAEYVQAOLOG
EGHLEVEFSPHPIITTSVEEMRRRAAGAGVGSILRRQDSTPRM:LEALGLAMOGP
VPWGRILFPAGGRVPLPTYPWOREYVTEADPAKSNAGDRGRVAGGHLIGEMOTLST

QSTRLEMTTLLDKRLPMLGDRHVGVGAVVPGAAVLEMAISSGAEALGDGPIQITDV
LAELAFADDAALVQVYTTQPSGRLOPOLASRAGAPASRVHARGALLVEETE
VPAGLTSVAVRARIQASMPAAATYAEITFMGLQYAPAGPAGIATLMGEGALGRVLP
DAAGSAEYRIRHPIALDADFOYVGSIFAGGERTPVVPEVGLRLIQRPSGLKCHA
RYVNHGRQTPDROGADFWYDSSGAVAAEVSGLVAVRLPGGVARRRDEDMFLELEPE
AAVGTAKVNAKRWMLLGGGGGLGALRLMSRGGAGVVAHAESNTDVRALAAAF
DSQAPVAVHLSLDGGELEDPLGAGALDAPRSADVANFELDLVRCGDSVLTAV
OLAGGFPDAPRLMLTTRGAOVAGADVSTQAPLLGGRVITAMHADLRCAKRVLD
PRPBGELAILLAELIADDAEVALRGSEKCVARIVRRQPTERRPGRJESCPVTVT
IRADSTLYTVTGLGGLSVACHLAERGAGHLYLVRSQSAVTEQAAVAALEKARAR
VYVAKADVADRAQLEKRLREVTTSMPLEVVHAAIILDDGLMOQTPAFRRVMAK
VOGALHIALHLEAPLSEFVLVAGVGLGSPGQVYAANFELDLAHHRKQGVAPA
LSYDMLFEAEVGAALAEEDRGALVSRGKMSLTDPDGLSALALLSGRQVQVMVNP
PLWVLEYLPAASSRMLSRITVAHRAASAGGAPAGDGLLRILAAEEDSARSGLLEPLR
AOISQVLRPEKRTIEVDAPITSLGMSNGLERMRNTEAMGCTTVPATLLMTPTVAA
LSGHLAREKCEAPVPSPTTADSAVEIEMSDDLTQLIAKFKALT

16134.37907
/gene="epod"
/locus="epod"
16134.37907
/note="type 1 polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTGPVTAQONPLKQAAIITIQLEERLAGLAQAELEPEPIAV
GIGCRPPGADAEAFWELDLDERDAVOPLDRMALVGAAPVAVHMGALLPEPIDC
PFAAFEGISPREARSLDPQHRLLEVAWELGELWAGIPRSIDSRGVGVATTADYA
RIVARLPREEDAVSATGNMLSTAAGRSLYTGCGCLTVDAAGSSSLVAILACRS
LRAGESDLAAGVSTLSPDMWEAAFTQALSPDRCTRPVDSAGSPVRECEGCIYV
LKRISDAQDGRINAILIGSAINHGRSTGLTAPVLAQVLYLREKALSAHYEAKAV
DVEYHGTSTLSDPLEVALRATVGPARDGRCVAGVKTNIIGLEAAAGVAGLIK
AALSTHERIPRLNRLTNPRITRLESALALATEVPPRPTRDRPREAVSSFGMGIT
NAHVLEAPAVELMPAEPARSAILVLSKSGALDQAARLREHLDHMPLEGLADV
AFSIAATRSAMTHRLAVATRSGLLAALSAVQOGTTPAGAAKCIASSSGKALIFT
GGGAGTGMGRGICAMPAPAFREAFDRCTLPREDLRLREVMAAGSASRSLIDGT
AFTOPALREYVALTRLMRSWGVPELVLGHSIGELVMAACVAGVSELECEVRLVAAK
RLMGGLSAGCAWSTIGAPAEVAAVAPHAAMVIAVAGPEQVVIAGVQAOALAA
GFAARGVRTKRLHVSHAFHSPLMEPLMEFGRVAAASVTYRRPVSLSVMSLCKVTDDE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 5, 6e-10;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCAGCGTCCGAGCGCGCCCTCGTGGCACCCCGCCGACGCGTGTGAT 60
|||||
Db 27200 CGCAGCGTCCGAGCGCGCCCTCGTGGCACCCCGCCGACGCGTGTGAT 27259
|||||

QY 61 CGCCGCGGTCAACGGTCTTACGCCCTCGTGCATGCTGGCG 101
|||||
Db 27260 CGCCGCGGTCAACGGTCTTACGCCCTCGTGCATGCTGGCG 27300
|||||

RESULT 2
LOCUS AR172664 71989 bp DNA linear PART 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien B., Katz, L., Khosla, C. and Tang, L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1.71989
Location/Qualifiers
/organism="unknown"
/protein_id="AAF62883.1"

BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN
Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 5, 3e-10;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCAGCGTCCGAGCGCGCGCCCTCGTGGCACCCCGCCGACGCGTGTGAT 60
|||||
Db 27200 CGCAGCGTCCGAGCGCGCGCCCTCGTGGCACCCCGCCGACGCGTGTGAT 27259
|||||

QY 61 CGCCGCGGTCAACGGTCTTACGCCCTCGTGCATGCTGGCG 101
|||||
Db 27260 CGCCGCGGTCAACGGTCTTACGCCCTCGTGCATGCTGGCG 27300
|||||

RESULT 3
LOCUS AF210843 68750 bp DNA linear BCF 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M.,
Nowak-Thompson, B., Engel, N., Toupet, C., Straumann, A., Cyr, D.D.,
Gorlach, J., Mayo, J.M., Hu, A., Goff, S., Schmid, J. and Lidgon, J.M.
TITLE The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
Chem. Biol. 7 (2), 97-109 (2000)

JOURNAL
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar, I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agricultural Research Institute, Inc., 3054 Cornwalls Rd, P.O.Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source 1.68750
Location/Qualifiers
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARVYLDGGRATERHVLAEARGLEDLRLAREH
LRTOGGSPHCWICGLDITVELLAHOPLASTISFHHARSLRHDDMSDMKXDPALY
RLILARAGAPPLREYEEERRRATQDEARLMLAAAPPCAPPLPREDDANGLPLGP
MSPEVAAEERLRAASYATPELACALALLAGTGWSGIPALEMLPENLLGFGLEPT
ALAAAGPISSEALRGARLFAWSVSVSSKSGOLNIDELMELRLTVRAMGANDN
LSFERAEALAAEVRRLRAOPAPFAGAGIAGVSSSGSLGTVDGALYSGDND
IYMFGRISPVVLAGTDFPELAPLSOMLEVAHANGTISKVLTGSPILVARN
QAPMSLIVHARGMAMVNOAMPDPERGAPVYVORSTIEFEHPTRCRHPHAGSAS
LACDEHLIVCELSAGRELKMRHPRHPRGAPSPFATLGEHPATATYPTSTLLNATV
WADPDRRAILGVDRKTVGEPTVLAETRHPPAHVSDRDIFFALTGPDSRDWHVEHIR
SGASTVADVOROLMDRPDVLNRLGFTTNDRIITLLARS"

CDS
1.71989
Location/Qualifiers
/organism="unknown"
/protein_id="AAF26905.1"

```

/db.xref="GI:6724319"
/translation="MGALISVNAFGCALGAEEREPQDDGAGLALAPREVMALFVN
AGMRPGLVAVAGDDVAVAVAGVETIGESAGMBRDPTRIASMTKATTPAAMVILV
EGRKIDISDPMDTRPELVNKKVAVATIDGIDETMBVAPRITVTRDLMTEFGISPA
SSSPORLIDELGVLNMOIVYMTFGDDEHMLRELOTLPLMHOGANMNYNTSIOCVL
VGRAPADGPAPURERIALFGKROTDEHYVADPKLARKAGCCTFLDEYKTRADD
GAEHAYVSPAFRFGSGAGLVSTIDOLLEFARNMLGAEGRKRLASVYKMTKADND
TPAOKAASSTPEFGFEHMGIGENAYVATDAVSEVPGATGMDGFGTSMINDGRELE
IGLVMTSAGLETSGLAERWMSVATPESA
3413..3536
"note="Ofcs; similar to Na/H and drug/H antiporters"
/codon_start=1
/translation_table=11
/product="unknown"
/protein_id="AA026106.1"
/db.xref="GI:6724210"
/translation="MHOLEROVQLSLVTALILVYARASGELARRLKPEPELGEFG
GVVGLSPVGLAVAGPRLALFEDPAGVAVLSGISWIGALLLILMANGIDVDGILRK
RPGALISVGLAIAPPLAGAFVLAFAFSLVATIDGIDETMBVAPRITVTRDLMTE
RRAVAQVTLADAVSEVAGVILVLTVPFLAALTORGLHPLGALFGLVLSNAPKPT
LHTLMAMKVAAGVSEVAGVILVLTVPFLAALTORGLHPLGALFGLVLSNAPKPT
NRLPLDVGQLELVAGLVPFPIEVLAVTGKRVVSOIRTPAAGVTALLATATKAPPA
SPALLILMERKAPDTEESARLEBEASARAVIPGERLIVYVHALPGFATDIESSE
IVASKRLEGGVETIDELSEEQAPGSPRAAGSAGLALGARLVGIMRORLEKRS
IOALILRASRSHDLVFLCANISAPARMSGRLODAIVORAEENVLYVMDRPPSRVAV
ARSLVIVDPEGRGLRGRVRSRVHVAHSDETRELARAPYDLVLICGYHGLGRLL
YLSGVSEVYVRSVPVALLVAGKREQVR"
/complement(5612..3592)
"note="Ofrc"
/codon_start=1
/translation_table=11
/product="unknown"
/protein_id="AA026507.1"
/db.xref="GI:6724241"
/translation="MKRIKIRTCALAEGVIGGSGNGPDSHEPACILNASDRDA
OVALTVFSPRDPAGPVRVYVAPRRHVRHFRNDLTERPEIPADTIDYASYIISDAP
OHTLDSROKADMLNSTLATITDNE
6228..6675
"note="Ofrs"
/codon_start=1
/translation_table=11
/product="unknown"
/protein_id="AA026918.1"
/db.xref="GI:6724241"
/translation="MKRVIVGREGFRIRIGHTIGLLASMLALACGGSPSEKTVOGTRLP
GADRARVADVDPPAAATTRLVDVVAVILSPERLEASERFVWQRPSPSPRRAGVLD
YNADSRKGIAETTPYANPELLITAEKOSSPQSPSAVIGTPIVSG
7610..11875
/gene="epoa"
/protein_id="AA026918.1"
/gene="epoa"
"note="Epos A"
/codon_start=1
/translation_table=11
/product="polyketide synthase"
/protein_id="AA026919.1"
/db.xref="GI:6724255"
/translation="MADPRIERAEADPIAVGASGRLPFGVIGDSGEWTLLEGSRDVI
GRVAFEDMAAENADDDPAZGCTPTVRSFISDVAFCDAFPGISPREALRMDPARH
LLEVCHEAENNAFNAIPNALIGENGVEFGISPEFSEVAPLQATASATIDHNGILGM
PSVNGATVYALIDRBCVAKNDVAGSSIVYVHALCOSLRSGEGTASGLVLSLSPG
STIVMSKRTALDRBCVAKNDVAGSSIVYVHALCOSLRSGEGTASGLVLSLSPG
SAIHNDASGGLVPMGSSOILKRALADACGASVGYIEAHBTGTTLDDPLEIDON
LNATVGLGRVATPLGSGVITLGPETASGITGLRVYLSLQGLPAHLHAKLQV
PRISMGRLVTRARPRMIMTPRKASVSSFGSGNHAHYIEAATACCPAPE
RPAVLVLSRTASALDAQRLRHDLETPQCGADVATSLATRSAMERLALVAAT
REAVDLELRFJNEEDLDRPVLVMMAEPASVDALIDOTAFOTPALETTEVLAALAMWS
WGVEPELVIAHSGEIVLACVAGVSELEDAVFLVVARGLMOALDPAGCAVSTIAPAR
LVAAVAPAAHAYSAISVNAIDGVYIAGAGGVVHAIAAAMARGARTKALHVAIFAG
LAPLAHLEAGVRAEISVYRPSITLVYSLSGKACTDEVSFGVWRHARVRFADCG

```

Query Match

96.8%; Score 97.8; DB 1; length 68750;

Best Local Similarity 98.0%; Pred. No. 2.1e-09;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 60
|||||
Db 32812 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 32871
|||||

Qy 61 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 101
|||||
Db 32872 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 32912
|||||

RESULT 4
LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 96.8%; Score 97.8; DB 6; Length 68750;
Best Local Similarity 98.0%; Pred. No. 2.1e-09;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 60
|||||
Db 32812 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 32871
|||||

Qy 61 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 101
|||||
Db 32872 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 32912
|||||

RESULT 5
LOCUS AR199551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 96.8%; Score 97.8; DB 6; Length 68750;
Best Local Similarity 98.0%; Pred. No. 2.1e-09;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 60
|||||
Db 32872 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 32912
|||||

Db 32812 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 32871
Qy 61 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 101
|||||
Db 32872 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 32912
|||||

RESULT 6
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 96.8%; Score 97.8; DB 6; Length 68750;
Best Local Similarity 98.0%; Pred. No. 2.1e-09;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 60
|||||
Db 32812 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 32871
|||||

Qy 61 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 101
|||||
Db 32872 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 32912
|||||

RESULT 7
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 96.8%; Score 97.8; DB 6; Length 68750;
Best Local Similarity 98.0%; Pred. No. 2.1e-09;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 60
|||||
Db 32812 CGCAGCGTCCGAGGCGCCAGAGTGGCCCTCCGTCGGACCCCGCCCGCCACAGGTGTGAT 32871
|||||

Qy 61 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 101
|||||
Db 32872 CGCCGCGGTCAACGCGTCTGACGCGCTGATCGTGGCG 32912
|||||

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothonones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 96.8%; Score 97.8; DB 6; Length 68750;
Best Local Similarity 98.0%; Pred. No. 2.1e-09;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 CGCAGCGTCGAGGCGGAGGTGGCGGCTCCGTGGACACCCACGCGCGCGGTGTGCAT 60
DB 32812 CGCAGCGTCGAGGCGGAGGTGGCGGCTCCGTGGACACCCACGCGCGCGGTGTGCAT 32871
OY 61 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 101
DB 32872 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 32912
RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epothonones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 96.8%; Score 97.8; DB 6; Length 68750;
Best Local Similarity 98.0%; Pred. No. 2.1e-09;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 CGCAGCGTCGAGGCGGAGGTGGCGGCTCCGTGGACACCCACGCGCGCGGTGTGCAT 60
DB 32812 CGCAGCGTCGAGGCGGAGGTGGCGGCTCCGTGGACACCCACGCGCGCGGTGTGCAT 32871
OY 61 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 101
DB 32872 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 32912
RESULT 10
LOCUS AX024384 16124 bp DNA linear BCT 15-SEP-2000

DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyanthum cellulosum.
ORGANISM Polyanthum cellulosum.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyanthum.
Bayer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source Location/Qualifiers
1..16124
/organism="Polyanthum cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 62.0%; Score 62.6; DB 1; Length 16124;
Best Local Similarity 76.2%; Pred. No. 0.0081;
Matches 77; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
OY 1 CGCAGCGTCGAGGCGGAGGTGGCGGCTCCGTGGACACCCACGCGCGCGGTGTGCAT 60
DB 1721 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 1780
OY 61 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 101
DB 1781 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 1821
RESULT 11
LOCUS AX024277 16124 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE Polyanthum cellulosum.
ORGANISM Polyanthum cellulosum.
REFERENCE 1 (bases 1 to 16124)
AUTHORS Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyanthum.
Bayer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source Location/Qualifiers
1..16124
/organism="Polyanthum cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 62.0%; Score 62.6; DB 6; Length 16124;
Best Local Similarity 76.2%; Pred. No. 0.0081;
Matches 77; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
OY 1 CGCAGCGTCGAGGCGGAGGTGGCGGCTCCGTGGACACCCACGCGCGCGGTGTGCAT 60
DB 1721 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 1780
OY 61 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 101
DB 1781 CGCGCGGTCAACGGGTCTGTGACGCCGCTCGTGGCTGTGGC 1821
RESULT 12
LOCUS SC167 26195 bp DNA linear BCT 12-MAY-2002
DEFINITION Streptomyces coelicolor cosmid 167.
ACCESSION AL591083
VERSION AL591083.2 GI:20520686

KEYWORDS	decarboxylase; diaminobutyrate-pyruvate aminotransferase; epoxide hydrolase; FAD-binding protein; integral membrane transport protein; oxidoreductase; regulatory protein; scot; secreted protein; type I polyketide synthase.
SOURCE	Streptomyces coelicolor A3(2).
ORGANISM	Streptomyces coelicolor A3(2).
REFERENCE	1 (bases 1 to 26195) Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
AUTHORS	Redenbach, M., Krieser, H.M., Denapate, D., Eichner, A., Cullum, J., Kinashli, H. and Hopwood, D.A.
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE	97000351
PUBMED	8843436
REFERENCE	2 (bases 1 to 26195) Seeger, K. and Harris, D.
AUTHORS	Unpublished
JOURNAL	3 (bases 1 to 26195)
REFERENCE	Cerdeno, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
AUTHORS	Direct Submission
TITLE	Submitted (01-MAY-2001) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
JOURNAL	CB10 15A E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT	On May 9, 2002 this sequence version replaced gi:13940311. Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/projects/S-coelicolor/) CDS are numbered using the following system eg SC/B7.01c. SC 1'S. coelicolor), 7b7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4781(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl . CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or atc) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 1g7 overlaps cosmid SC1c4 and cosmid SCBAC8D1.
FEATURES	Location/Qualifiers
SOURCE	1. 26195 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid 1g7"
misc_feature	2. 118 /note="nominal overlap with Streptomyces ccelicolor cosmid SC2c4"
misc_feature	complement(4..618) /note="Pfam match to entry PF00698 Acyl_transf, Acyl transferase domain, score 5.10, E-value 2.8e-10"
misc_feature	complement(928..1437)
misc_feature	/note="Pfam match to entry PF02801 ketoacyl-synt C, Beta-ketoacyl synthase, score 346.10, E-value 3.8e-100" complement(1459..2214) /note="Pfam match to entry PF00109 ketoacyl-synt, Beta-ketoacyl synthase, score 469.40, E-value 3e-137" complement(1672..1722) /note="PS00606 Beta-ketoacyl synthases active site" complement(2284..2487) /note="Pfam match to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 66.20, E-value 1.1e-16" complement(2362..2409) /note="PS00012 Phosphopantetheine attachment site" complement(4861..5829) /note="Pfam match to entry PF00698 Acyl_transf, Acyl transferase domain, score 307.60, E-value 1.4e-88" complement(6142..6651) /note="Pfam match to entry PF02801 ketoacyl-synt C, Beta-ketoacyl synthase, score 328.60, E-value 7.1e-95" complement(6673..7425) /note="Pfam match to entry PF00109 ketoacyl-synt, Beta-ketoacyl synthase, score 458.60, E-value 5.1e-134" complement(7495..7698) /note="Pfam match to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 55.10, E-value 1.3e-13" complement(7909..8832) /note="Pfam match to entry PF00698 Acyl_transf, Acyl transferase domain, score 280.60, E-value 2e-80" complement(9139..9654) /note="Pfam match to entry PF02801 ketoacyl-synt C, Beta-ketoacyl synthase, score 274.60, E-value 1.3e-78" complement(9676..10416) /note="Pfam match to entry PF00109 ketoacyl-synt, Beta-ketoacyl synthase, score 353.40, E-value 2.5e-102" 10776..10781 10793..12115 /gene="SC1g7.02" /note="SC06276" 10793..12115 /gene="SC1g7.02" /note="SC1g7.02", possible secreted protein, len: 440 aa; similar to TR:Q9L0A8 (EMBL:AL163003) Streptomyces coelicolor putative secreted protein SC2c4.09, 468 aa; fasta scores: opt: 777 Z-score: 867.4 bits: 169.7 E(): 1e-40; 36.303% identity in 449 aa overlap. Contains Pfam match to entry PF01360 Monooxygenase, Monooxygenase" /codon_start=1 /transl_table=1 /product="putative secreted protein" /protein_id="CAC37877.1" /db_xref="GI:13940313" /db_xref="SPRMBL:Q93613" /translation="MRRKREGRAIVLGSMAGLFSARALSETFANTVTVDRDELTCFN ELRGGLPGHNLGLRGGQITPEYRPGITTEORBARAGTQADVADRVNTINKRLA KYSGGLATSSRPELRHNRDYLALPEYSLERAVOSTLPTADSRVNGVYVRA DGEETIHADLVDTGRGRTSWLOEFGTEAVOEKKNITGLYTRRYRVDEARE GNSISINVASAGVPRGAIQKIDSDRAIVRAYGLIGHPPTDPEGIGFATKSLAADPI HEVLTVESEPIDDPVAKFPTNLRHYORMDPFGILVIDAVCSYVAAOGATVSA LGSVLVRLHADTHRPDLRYFEDLADAVGCEMAYVGDLSPEVEGGRFPEVQRA HAFIAVQVEATRDPAVAARAVSRVIGLVDSPAVLHDSPTFAIAGAQR" 11294..11908 /gene="SC1g7.02" /note="SC1g7.02" /note="Pfam match to entry PF01360 Monooxygenase, Monooxygenase, score 4.10, E-value 3.3e-05" 12101..12104 /gene="SC1g7.02" 12112..12972 /gene="SC1g7.03" /note="SC06277" 12112..12972 /gene="SC1g7.03" /note="SC1g7.03, possible epoxide hydrolase, len: 286 aa;

similar to TR:043566 (EMBL:D16628) Arabidopsis thaliana epoxide hydrolase (EC 3.3.2.3) Seh or F1848.11, 321 aa; fasta scores: opt: 351 z-score: 406.4 bits: 83.3 E(1); 4.8e-15; 29.560% identity in 318 aa overlap. Contains Pfam match to entry Pf00561 abhydrolase, alpha/beta hydrolase fold"

/codon_start=1
/transl_table=11
/product="putative epoxide hydrolase"
/protein_id="CAC37878.1"
/db_xref="GI:13947514"
/db_xref="SPTREMBL:Q93S12"
/translation="MTA:RIGDLGVFDVDVTPDDEPVLILHGFPHNKESWTETAP LHAAGLTVAPDQRY:IPDAPRAVADRIPLHDLADAGVLDALGVSSAHVGDHWG AAVAYLAAHAPGRVY:ITALA:PHLDAYQAHVQVDEQHSQVEFLVSGADHF LADGAOLHAPETQAGE:VLTDQA:IAVAVHTAAGTDLALNWRANRLDGPDPFG PVTVPPTTFTVSKRTDPAV:ELAVAKTEEVYTGPRLLTLEKTSKHWOPDDPTVAEIL ARVAVROTH"
12262..12948
/gene="SCI1G7.04"
/note="Pfam match to entry Pf00561 abhydrolase, alpha/beta hydrolase fold, score 74.20, E-value 2.7e-18"

RBS
gene
CDS
13015..14646
/gene="SCI1G7.04"
/note="SC06278"
13015..14646
/gene="SCI1G7.04"
/note="SCI1G7.04"
protein, len: 543 aa; similar to TR:Q9RM9P (EMBL:AF201913) Streptomyces globiaporus transmembrane membrane efflux protein Sgcb, 521 aa; fasta scores: opt: 1451 z-score: 1413.5 bits: 271.2 E(1); 3.8e-71; 46.243% identity in 519 aa overlap. Contains possible hydrophobic membrane spanning regions"

/codon_start=1
/transl_table=11
/product="putative integral membrane transport protein"
/protein_id="CAC37879.1"
/db_xref="GI:13940315"
/db_xref="SPTREMBL:Q93S11"
/translation="MSSMTAPAEVELRACPREWEGLLALLPVTMTADLVMTATPY LTADLPTSSQLMTTDLGFEWTCGLVVMGTLDGRIRRLIIGSLGVIAVLAIA YSTSEMLIVARLALGVAGAAVLPSTLSLIHMFDDORARATATWTALSVGAIG PVIGVLLLEHMMGVSFLM3VVMVLPVLLPEKDPGAGRLDLSAVLFAAI LPVVGIRKFAHEHWSLANLCAIVIGAAFTVVRROSLFPLDMLFTRFTGA LUTLEGMALNGVEYLVF JYLLVAGELTLPAGLMLPLGAGLIGSOLPVLAKRE RPAVYITAGLVVTLIGFNL IAVAGPDDSGIYVPAAGLAVIENGVAPIIGLTAAGA APERKAGAAATGATAYDULAFGIATVGSVAIVAYRSIGIADTAPAGIPAAEFARL TVGAKAKAESLPDVEBQJLTAAREAFETSGPHATVAVSAGMAVLTAVVALVLRHIP

Query Match 51.7% Score 52.2; DB 1; Length 26195;
Best Local Similarity 71.1% Pred. No. 0.57; Mismatches 28; Indels 0; Gaps 0;
Matches 69; Conservative 0;

QY 5 GCGTCGAGCGGAGGTGCGCTCCGTCGACCCAGCGCGCGCGTGCATCGCCG 64
DB 204 GCGTCGAGCGGAGGTGCGCTCCGTCGACCCAGCGCGCGCGTGCATCGCC 145
QY 65 GCGGTCAACGGTCTGACCGCGCTCGATCGCTGGCG 101
DB 144 GGGGTCAACGGCGCCGAGTCGTGTGATCTCGGGG 108

RESULT 13
AX211706 AX211706 27541 bp DNA linear PAT 06-SEP-2001
LOCUS
DEFINITION Sequence 2 from Patent W00159126.
ACCESSION AX211706
VERSION AX211706.1 GI:15523938
KEYWORDS
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei.

REFERENCE
AUTHORS
1 (bases 1 to 27541)
Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Strom,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Guilken,O.M.

TITLE
Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility

JOURNAL
Patent: WO 0159126-A 2 16-AUG-2001;
Norges Teknisk-Naturvitenskapelige Universitet (NO); STRYPLSEN
IND OG TEKNIISK FORSKNING VID NORGES TEKNIISKE HOGSKOLE (NO);
ALFHARNA AS (NO); SINVENT AS (NO); ZOTCHEV, Sergey Borisovich
(NO); SEKUROVA, Olga Nikolayivna (NO); FJAERVIK, Espen (NO);
BRAUTASET, Trygve (NO); STROM, Arne Reidar (NO); VALLA, Svein
(NO)

FEATURES
source
1..27541
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"

BASE COUNT 3517 a 10766 c 9529 g 3729 t
ORIGIN

Query Match 50.1% Score 50.6; DB 6; Length 27541;
Best Local Similarity 70.1% Pred. No. 1.1; Mismatches 29; Indels 0; Gaps 0;
Matches 68; Conservative 0;

QY 5 GCGTCGAGCGGAGGTGCGCTCCGTCGACCCAGCGCGCGCGTGCATCGCC 64
DB 13766 GCGATCCAGCGCACGACGAGACAGATCGGGCGCACTCGACAGACGCTGGCATCGCC 13825
QY 65 GCGGTCAACGGTCTGACCGCGCTCGATCGCTGGCG 101
DB 13826 GCGGTCAACGGCGCGAGTCGTGTGATCTCGGGG 13862

RESULT 14
AF263912 AF263912 123580 bp DNA linear BCT 24-MAY-2000
LOCUS
DEFINITION Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
complete sequence.
ACCESSION AF263912
VERSION AF263912.1 GI:8050835
KEYWORDS
SOURCE
ORGANISM

Streptomyces noursei.
Streptomyces noursei

REFERENCE
AUTHORS
1 (bases 1 to 123580)
Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Ström,A.R.,
Valla,S. and Zotchev,S.B.

TITLE
Biosynthesis of the polyene antifungal antibiotic nystatin in
Streptomyces noursei ATCC 11455: analysis of the gene cluster and
deduction of the biosynthetic pathway

JOURNAL
Chem. Biol. 7 (6), 395-403 (2000)
PUBMED
10873841
2 (bases 1 to 123580)
Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
Valla,S. and Zotchev,S.B.
TITLE
Direct Submission
Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
N-7489, Norway

FEATURES
source
1..123580
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
/gene="nysF"
complement(46..783)
/gene="nysF"

```
/function="presumably post-translationally modifies ACP
domains on PKS"
/note="putative 4'-phosphopantetheine transferase"
/transl_start=1
/transl_table=11
/product="Nysf"
/protein_id="AAE71762.1"
/db_xref="GI:8050836"
/translation="MIELLPATVATPAAYADDBRPDGRLLSSEREVIYAAVESROE
FTVVRRLARALRLGHPDRAIIPNRGAQONPPGYGSMTRCAGYRAAASPAESLA
ASVIDAENCPDPAGVNLATLPSERPHILAAHDPVHMDDLSEAKSVREAMP
LTQRELDSEAEIADPTQGAFTARLLPVPGLLGRRVTEPGRMHSTPALTTVAHL
PAPTPRDRERHRLTVNSDLPPTPG"
/compelment(867. .2684)
/gene="nysg"
/compelment(867. .2684)
/gene="nysg"
/compelment(867. .2684)
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="Nysg"
/protein_id="AAE71763.1"
/db_xref="GI:8050837"
/translation="MASPDLEERTAPRPARRVLGLRPRRSVALAVMGVGIYL
NAPGLLGRVLDLADGVGVGPAPGIDFAIGRLVLLALVYVASLPMAGOR
LVASAVWRTHLDRDAEKLRLRHPDRPAGELSTRTNDIMLOOTLTAE
LITISLTMLVIMLVISLAVNMLISPVSAIARIKRAQPHYAOWSMNGTL
NAHEEVCTGHALIKGDRRAAEERDADCNDVYRAAQAQFASGMEYVNAVIL
GVAVLVIGAMKVINGTLLGDVQALVAROSOPIVETASVAGLQSGIASORVF
TLIDAEQOAPDLRPGTPARAEGRVEFTDVSFSPDTPIENLSLVEEPTVAIG
PTGAKTTLGNLMRFYEPEDSGRLDDGTPTMTDRDLRSRGLVLODMVFGTIA
ENIAGAPGACRADIEEARAATCADRPITRLPGQYDVLDESGTASGEKOLLTVR
AETARAVLYLDEATSSVDTRTEVLIORANSLRAGTSFYIAHLSITDADLIYVM
DAGRIVEQGHDLCAQGLYARLHAARTHTPAGAAG"
/compelment(2662. .4416)
/gene="nysH"
/compelment(2662. .4416)
/gene="nysH"
/compelment(2662. .4416)
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysH"
/protein_id="AAE71764.1"
/db_xref="GI:8050838"
/translation="MLRLRLAOLRPYAMATALVALQVLIIGTLPLTGAALIDQ
GVVRCGGGRITELGVVGVVALVOIQAALGAALAAATATAMGRDLSALFRILDS
ARIGRRGPTSLTRSYNDVOYONLAOTGFIYVCAPMLCISVILLALRODPLAL
LVALLVAVACFGLLARMGTLYARQOLTLDRGLRLREAITGVRRVRSFRRDHERA
RFAOTNDAPLVSRRYCRILATMLPVILLMNGFTYVALMTGSHRIDAGRMPIGSLSA
LSTLSLIMSVMALFVPLSVPRARCAGRIAEVLDIGSSVAFPAAPQVPRGAGTI
ELCAAGRYRPGAEPVLRVDVLFVEPERIAVIGSGSKTTLNLTIRLADITEGAV
RNGCTVRELTAATLAAVGVFORPRLTFSVGLSALNRFGRPALTDEBELAALVAA
ADVVARMPDLDAEITOGGNGSGGROKORLIRALLRREIYELFDDCFALDOATA
ALHTALVPYTAGATVITVQRIASGRADRIIVLDRGRVAAQGHDLTLTSTPYTRI
ALSQLTBEERAHGLAGR"
/4714. .5748
/gene="nysDIIT"
/4714. .5748
/gene="nysDIIT"
/function="presumably involved in mycosamine biosynthesis"
/note="putative dGDP-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NysDIIT"
/protein_id="AAE71765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITGODGSYLAEBHLSQGYVWGLIRGCANPRKSY
SRLASLDFITDGLMDQGLVSAVDVQDPDEVNLGAIISVPMSWDAELVTEWNGG
VLRMLERIRWVSGLSRIVSPRGQIRFYQASSSEMGALETIPORETTLFHPRSYPG
AAKAYGHYITRNRSESGVAVSGMLFNEHSPRGGEFYRKISLAVARIKQGLQDLT
ALGNLDVAVRDMGYAGDYVRAMHMLQDQDADGVYIGTGQHSVADAVRIAEHVGLMW
```

```
EDYVIDPDLVPRAEVEVLCADSAKQDRGLMKRPDVEPTLMRMVDSDLAQVRENO
YGDVILAAAN"
5930. .34363
/gene="nysI"
5930. .34363
/gene="nysI"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="NysI"
/protein_id="AAE71766.1"
/db_xref="GI:8050840"
/translation="MNEOKRLDYKIKATADLRTRRVRVKKLSAQAEPVATIGMCR
YGCGRSPEDLMKRVENGCVTFPPDPDRKMDLEALAAATPASGGTTHDAPDDADPF
GISPREVNAADPOQRVVLSYSEAMEAFEEAGVIDPSVKSGRTGVETGAADQYRGPAD
AEGFOLGNTGSVLSGRISYTFGVGAVAVVDACSSSLVAVHLATOLARAGCTLAL
AGGVYIMSGGTETEMROGGLSADRCRSGTADGTGNAEGGILYELRLSDAVRN
GHETLAVRGTAVNQDGSNGLTPNPGRSQOQYVQOALVNAALAGVLDVVEAGTGT
TIGDPVEAKOALLNTYGNRRADRPILIGYSKNSLHQAAAGVAGVYKMMARMHGTL
PRTLHAEPETHHNDMSQAVRLLDITDMPATGAPRAAVSFGISGTNAHTITBEQAP
EPOEDATATQDDAAGSTPATAPVPVGVVLLSGRPDALRQOAAALRALDTPRRP
DLIDLASHLATTRAGFEHRAVLTLATDHPALTDIGLTLADADDPAAPAWITGTTRAT
RLAVLFTGQGAQRGLAGRELAARPAPATLADALDAFTPHLDRPLEVLMGDDAAL
DRYVLAQPALEFVAVALYRLIESFVRPHDLAGVSGETIYVAHLAGVLSADATVLA
ARGLMALPDGGMATVQVASEADVAPILAGHEDOVAIAAVNPSANVLSGAEVTVLA
LABOLADGKTRRLRYSHAFHSTLMEPMLDARAVVEDITLOPPLIPVSNLGRKA
TVQOLTSADVWVHVRHVRPAQIDMLANHDPTAFLELPDGLVLSMADCDLADA
DAVLTPLRGRREHRLTTLALGHHVAGTLDWTGCFAGTGARNDLPTVAPQRXY
WPKALQGLADLSVSGAAHHPILSAVSLADAGTLLTGRSLRQHPMLDVRVGT
TTLPGSTAFELIYVAGDEVGCDRVEELTLAPLLPEOGVOYOLMIGPDVSGRT
VNHARBDTGDOPMTNAATGVTTLTADASQOLPASSSGOCTPLAGDPHPLDAAQMP
AGAPPLDGHYDLADGFGYGFVFGGLRAMGCGVYAEVLEPRAGSDABEGLT
HPALDLALHAAPPTGIGERGGLPFSWBGVSLHAGCEGFEFGFMSPRRLALTY
ADGTCAPVLSVDSILVRSVATQOQLDAAARALFELDWTPOVPTADDPVALLTY
DPGLTLHAGCFADAPADPDLAALAAAGAPVPTVYLSLACTGDADPARSARHCAE
ALAAVOTMLDHEHEFAARLVETVTRGATVGRDVAAYAAVMGLVNSAGEENGCFALVD
DPGACGAALVAVALSVEPOLAVRGDLVAVRLVRLPTEVAGACGACGDDGDSG
VSRSGEAVLYTGGTGIGLAVLARHLVAEVCYRDLILVSRSGERAVOAGCLVAYELAY
GARVRVACVYTDAAVVELVGGHAAVSAVHAAGVLDGNGVALTGRIRLSAELPKYD
AVMHLHEATRGDLDAFVNFVSLAGVSAVQVNAANAFLMALMRRAREAGLPGSL
LWMPMEQSGMGTGLTDVDAERLARSVGPPLSAQGLALFDDAAGTDAVCPVRDL
LPVLRARGVEPPLRLSLIRVARRAAVAGSTAGVLAORLRLRDEGDRDMVLDIVYG
OVALYIGHATGCGVDARARPDGFSLTVELRNRLNTYGTIRLPATVYEDVTVPH
LATVDELGLTDEAVTQVPAAYAAVADVTYGNACRTPGVSSPEDLMRVLTBET
DAVSGFTPNKGWDESLYHPDPHPGTSYSTRSGGLEHAGEFDFGFGMSPRRLATD
SOORLLLESWEALIERAGIDPVSIRKSGTGVFGVMSYDSAMLASPEFGFGSGSS
PSLASGVAATLLEGPAVTVDFACSSLIYAMHMAOMALRSCEGLALAGGVYMSRP
AVYDFAROGTISPDGCKAPFAADAGVSGSEGVYVLERODAVNGHEITLAVNG
SAYNQDASNGITRAPNPSOQRYTROLASGGLTAGVDVYVEAHGCTTITGDPLEAD
LNAVYGRDRPERPLLLGSYSNIGHQOAAAGAVIKVLMVHMGVPRTLVADAS
SHWMSGAVELLSEQAMPEGTQVRRASVSFGISCTENVLYVEOPAGKAAIAAQA
ARTPGAPVPLLVSALRGAARLALGILQALRPDLDVALSLATTSRREORA
VVADROULLASGLALADRPDPVAVGEAGSKRPVLTGTGOSQOAAANGREHVEQ
PERRAADAIVACVAPDPLDRPLREPVVAEAGSEALLDTEGOSQOAAANGREHVEQ
VESNGVPRDPVAGHSIGETIAAHAVAGVLTLEDRCUYAARATLMQALPTGGMATIDA
TEDEIAHLDDTVAAIAAVNGPQSVISGDEAEATTIATPBERKTKRLVSHAFS
PRMDGMIDARIVAELIYVAPRTPVLSDDTRADAECVTEAYVRRHNRVAVRPAD
CVRLTRDAGATTELEISDGLITLMAADPTIGDHDDDELVPLMRAGREAEALATLAR
LOYGVGVDMAAVYLAGGARRTDLPYTAFOAHADHYWVLPPTPAALAAADPAOOLMA
VERGDARELADITIGGEBODLTPDISLILPATLSMRNGOKEHLDTLRYREVMRLSEP
TAPVLDCTWLLVASDRTANQPALDGLDALALSHGRVRLLLDSDCADRATVAREL
ARTADVDAATQVLSVPLDERDADDCPTLRGTLATVALVQALADDTGTAQGRVATRE
AVSTNPADVPVTHPQOAAAGLGRGVLALEHRLNGGLVDLQVDFERGQRLAGILAK
DAPDGEDOVALLRATGVSGRLVTRHTVLAETPTAAEFTATGTVLTGGGCGAGAEARML
ARAGAOHLVLTSRGPDAPGAELRABLEXGVSVAHHDLLADLDRDLALVFLVFSIT
LPLTLGVNHTAGVGHYGPDLPTSTAEFGPAAKATAGAAHCDLLADLDRDLALVFLVFSIT
AGVWGSNGSAYCANAYLDALALHRRARGLIATSVAKGMAARGAADAAYSETLKR
OQGLGLDPAAMTELTRAUVRODVTVAVVDQWORALETTSAPSLMLIGLEBEVRL
AABERETVAGTESSEVTRVRALEBPQLLDLVAETLDESATVYGHSSADAVEGRAR
RDVGFDSTLTVELERKIRGAATGLSPTWVFDVPTLELAQYRAELTAGVLEAVAGY
ATGAGDDEPITAIIGMACRFPFGVSSPQGLMDVLAGSDAISERPVPVNGQTLHLPDP
```



```
FT /tag- d
FT /note- "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc-RNA
FT 3621..4661
FT /tag- e
FT /note- "encodes acyl transferase (AT) of the loading
FT domain"
FT misc-RNA
FT 4917..5810
FT /tag- f
FT /note- "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole holey"
FT misc-RNA
FT 5856..6155
FT /tag- g
FT /note- "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS
FT 6260..10493
FT /tag- h
FT /label- epob_gene
FT /note- "encodes module 1, the NRPS module"
FT misc-RNA
FT 2031..3548
FT /tag- i
FT /note- "encodes condensation domain C2 of the NRPS
FT module"
FT misc-RNA
FT 2031..3548
FT /tag- j
FT /note- "encodes condensation domain C2 of the NRPS
FT module"
FT misc-RNA
FT 6861..6887
FT /tag- k
FT /note- "encodes heptacyclization signature sequence"
FT misc-RNA
FT 6861..6887
FT /tag- l
FT /note- "encodes condensation domain C4 of the NRPS
FT module"
FT misc-RNA
FT 7358..7366
FT /tag- m
FT /note- "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc-RNA
FT 7898..7921
FT /tag- n
FT /note- "encodes adenylation domain A1 of the NRPS module"
FT misc-RNA
FT 7898..7921
FT /tag- o
FT /note- "encodes adenylation domain A1 of the NRPS module"
FT misc-RNA
FT 8261..8308
FT /tag- p
FT /note- "encodes adenylation domain A3 of the NRPS module"
FT misc-RNA
FT 8411..8422
FT /tag- q
FT /note- "encodes adenylation domain A4 of the NRPS module"
FT misc-RNA
FT 8861..8905
FT /tag- r
FT /note- "encodes adenylation domain A6 of the NRPS module"
FT misc-RNA
FT 8966..8983
FT /tag- s
FT /note- "encodes adenylation domain A7 of the NRPS module"
FT misc-RNA
FT 9090..9179
FT /tag- t
FT /note- "encodes adenylation domain A8 of the NRPS module"
FT misc-RNA
FT 9183..9992
FT /tag- u
FT /note- "encodes oxidation region for forming thiazole"
FT misc-RNA
FT 10121..10138
FT /tag- v
FT /note- "encodes adenylation domain A10 of the NRPS
FT module"
FT misc-RNA
FT 10261..10306
FT /tag- w
FT /note- "encodes thioation domain (PCP) of the NRPS
FT module"
FT CDS
FT 10639..16137
FT /tag- x
```

```
FT /label- epoc_gene
FT /note- "encodes module 2"
FT misc-RNA
FT 10656..12033
FT /tag- y
FT /note- "encodes KS2, the KS domain of module 2"
FT misc-RNA
FT 12250..13287
FT /tag- z
FT /note- "encodes AT2, the AT domain of module 2"
FT misc-RNA
FT 13327..13899
FT /tag- aa
FT /note- "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc-RNA
FT 14962..15756
FT /tag- ab
FT /note- "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc-RNA
FT 15763..16008
FT /tag- ac
FT /note- "encodes ACP2, the ACP domain of module 2"
FT CDS
FT 16134..37907
FT /tag- ad
FT /label- epod_gene
FT /note- "encodes modules 3-6"
FT misc-RNA
FT 16425..17606
FT /tag- ae
FT /note- "encodes KS3"
FT misc-RNA
FT 17817..18857
FT /tag- af
FT /note- "encodes AT3"
FT misc-RNA
FT 19581..20396
FT /tag- ag
FT /note- "encodes KR3"
FT misc-RNA
FT 20428..20642
FT /tag- ah
FT /note- "encodes ACP3"
FT misc-RNA
FT 20706..22082
FT /tag- ai
FT /note- "encodes KS4"
FT misc-RNA
FT 22296..23336
FT /tag- aj
FT /note- "encodes AT4"
FT misc-RNA
FT 24069..24647
FT /tag- ak
FT /note- "encodes KR4"
FT misc-RNA
FT 24867..25151
FT /tag- al
FT /note- "encodes ACP4"
FT misc-RNA
FT 25203..26576
FT /tag- am
FT /note- "encodes KS5"
FT misc-RNA
FT 26793..27883
FT /tag- an
FT /note- "encodes AT5"
FT misc-RNA
FT 27966..28574
FT /tag- ao
FT /note- "encodes DH5"
FT misc-RNA
FT 29433..30287
FT /tag- ap
FT /note- "encodes ER5"
FT misc-RNA
FT 30321..30869
FT /tag- aq
FT /note- "encodes KR5"
FT misc-RNA
FT 31077..31373
FT /tag- ar
FT /note- "encodes ACP5"
FT misc-RNA
FT 31440..32807
FT /tag- as
FT /note- "encodes KS6"
FT misc-RNA
FT 33018..34067
FT /tag- at
FT /note- "encodes AT6"
FT misc-RNA
FT 34107..34676
FT /tag- au
```



```
FT FT /note= "encodes DH6"
FT FT misc_RNA 35760..36641
FT FT /tag= av
FT FT /note= "encodes ER6"
FT FT misc_RNA 36705..37256
FT FT /tag= av
FT FT /note= "encodes KR6"
FT FT misc_RNA 37470..37769
FT FT /tag= ax
FT FT /note= "encodes ACP6"
FT FT CDS 37912..49308
FT FT /tag= ay
FT FT /label= epof_gene
FT FT /note= "encodes modules 7 and 8"
FT FT misc_RNA 38014..39375
FT FT /tag= az
FT FT /note= "encodes KS7"
FT FT misc_RNA 39589..40626
FT FT /tag= ba
FT FT /note= "encodes AT7"
FT FT misc_RNA 41341..41922
FT FT /tag= bd
FT FT /note= "encodes KR7"
FT FT misc_RNA 42181..42423

Query Match 100.0%; Score 101; DB 21; Length 71983;
Best Local Similarity 100.0%; Pred. No. 2e-15; Mismatches 0; Gaps 0;
Matches 101; Conservative 0; Indels 0;

QY 1 CGCAGCGTCGAGCCGAGGTGGCGCCCTCCGTGGCACCCCGCCGCGCATGTCGAT 60
Db 27200 CGCAGCGTCGAGCCGAGGTGGCGCCCTCCGTGGCACCCCGCCGCGCATGTCGAT 27259
QY 61 CGCCGCGGTCAAGCGTCTTACGCCCTGTCGATCGTCGGC 101
Db 27260 CGCCGCGGTCAAGCGTCTTACGCCCTGTCGATCGTCGGC 27300

RESULT 2
AA255887
ID AA255887 standard; DNA: 68750 BP.
AC AA255887;
XX 10-APR-2000 (first entry)
DE Sorangium cellulosum 68.75 kb contig.
XX Epithlone biosynthesis; type I polyketide synthase; taxol substitute;
KM anticancer; ds.
XX Sorangium cellulosum.
OS
XX
XX
XX Key Location/Qualifiers
XX CDS 1..1826
XX /tag= a
XX /partial
XX /product= "Partial Orf 1 protein (AAV58580)"
XX /note= "No initiation codon given in the specification"
XX CDS /tag= b
XX /product= "Orf 2 protein (AAV58581)"
XX 3415..5556
XX /tag= c
XX /product= "Orf 3 protein (AAV58582)"
XX complement (5612..5992)
XX /tag= d
XX /product= "Orf 4 protein (AAV58583)"
XX 6226..6675
XX /tag= e
XX /product= "Orf 5 protein (AAV58584)"
XX 7610..11875
XX /tag= f

FT FT /product= "Type I polyketide synthase, EPOS A
FT FT (AAV58573)"
FT FT 11872..116104
FT FT /tag= g
FT FT /product= "Non-ribosomal peptide synthetase, EPOS P
FT FT (AAV58574)"
FT FT 16251..21749
FT FT /tag= h
FT FT /product= "Type I polyketide synthase, EPOS B
FT FT (AAV58575)"
FT FT 21746..43519
FT FT /tag= i
FT FT /product= "Type I polyketide synthase, EPOS C
FT FT (AAV58576)"
FT FT 43524..54920
FT FT /tag= j
FT FT /product= "Type I polyketide synthase, EPOS D
FT FT (AAV58577)"
FT FT 54935..62254
FT FT /tag= k
FT FT /product= "Type I polyketide synthase, EPOS E
FT FT (AAV58578)"
FT FT 62369..63628
FT FT /tag= l
FT FT /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT FT (AAV58579)"
FT FT 63779..64333
FT FT /tag= m
FT FT /product= "Orf 6 protein (AAV58585)"
FT FT complement (63853..64290)
FT FT /tag= n
FT FT /product= "Orf 7 protein (AAV58586)"
FT FT 64363..64920
FT FT /tag= o
FT FT /product= "Orf 8 protein (AAV58587)"
FT FT complement (64287..64727)
FT FT /tag= p
FT FT /product= "Orf 9 protein (AAV58588)"
FT FT 65063..65767
FT FT /tag= q
FT FT /product= "Orf 10 protein"
FT FT complement (65008..65874)
FT FT /tag= r
FT FT /product= "Orf 11 protein (AAV58590)"
FT FT complement (65871..66338)
FT FT /tag= s
FT FT /product= "Orf 12 protein (AAV58591)"
FT FT 66667..67137
FT FT /tag= t
FT FT /product= "Orf 13 protein (AAV58592)"
FT FT 67334..68251
FT FT /tag= u
FT FT /product= "Orf 14 protein (AAV58593)"
FT FT 68346..68750
FT FT /partial
FT FT /tag= v
FT FT /product= "Partial Orf 15 protein (AAV58594)"
FT FT /note= "No termination codon given in the specification"
FT FT WO9966028-A2.
FT FT 23-DEC-1999.
FT FT 16-JUN-1999; 99WO-EP04171.
FT FT 18-JUN-1998; 98US-0099504.
FT FT 24-SEP-1998; 98US-0101631.
FT FT 05-FEB-1999; 99US-0118906.
FT FT (NOVS ) NOVARTIS AG.
FT FT (NOVS ) NOVARTIS-ERFINDUNGEN VERW GBS MBH.
FT FT Schupp T, Ligón JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```


XX	29-NOV-2001	(first entry)
DT		
XX	Streptomycetes noursei nystatin PKS gene cluster DNA.	
DE		
XX	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;	
KW	antifungal; antibiologic; ds.	
KW		
XX	Streptomycetes noursei.	
OS		
XX		
XX	Key	Location/Qualifiers
PH	CDS	6337..34771
FT		/*tag= a
FT		/product= "NysI complete protein"
FT	CDS	34792..51099
FT		/*tag= b
FT		/product= "NysJ protein"
FT	CDS	51155..57355
FT		/*tag= c
FT		/product= "NysK protein"
FT	CDS	57503..58687
FT		/*tag= d
FT		/product= "NysL protein"
FT	CDS	complement (58786..58980)
FT		/*tag= e
FT		/product= "NysM protein"
FT	CDS	/note= "CDS does not include start codon"
FT		complement (59045..60241)
FT		/*tag= f
FT		/product= "NysN protein"
FT	CDS	/note= "CDS does not include start codon"
FT		complement (60238..61296)
FT		/*tag= g
FT		/product= "NysD2 complete protein"
FT	CDS	120628..121308
FT		/*tag= h
FT		/product= "NysR4 (long) protein"
FT		
XX	WO200159126-A2.	
PN		
XX	16-AUG-2001.	
PD		
XX	08-FEB-2001; 2001WO-GB00509.	
PF		
XX	08-FEB-2000; 2000GB-0002840.	
PR	10-APR-2000; 2000GB-0008786.	
PR	14-APR-2000; 2000GB-0009387.	
XX		
XX	(UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIG.	
PA	(SMTF) SINTER STIFTELSEN IND TEK FORSK.	
PA	(ALPH-) ALPHARMA AS.	
PA	(SINV-) SINVENT AS.	
PA	(DZIE/) DZIELEWSKA H.	
PA	(ZOTC/) ZOTCHEV S B.	
PA	(SEKU/) SEKUROVA O N.	
PA	(FJAE/) FJAEERVIK E.	
PA	(BRAU/) BRAUTASET T.	
PA	(STRO/) STROM A R.	
XX		
XX	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;	
PI	Valla S, Ellingsen TE, Sletta H, Gulliksen O;	
XX	WPI: 2001-557614/62.	
XX	P-BSDR: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,	
DR	AAE10149, AAE10150.	
XX		
XX	New nystatin polyketide synthase polynucleotides and polypeptides,	
PT	useful as antibiotics and antifungals -	
XX		
XX	Claim 1; Page 188-254; 266pp; English.	
XX		
CC	The present invention relates to the cloning and sequencing of the gene	
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme	

Query Match	50.1%;	Score 50.6;	DB 22;	Length 125401;
Best Local Similarity	70.1%;	Pred. No. 0.0013;		
Matches 68;	Conservative 0;	Mismatches 29;	Indels 0;	Gaps 0;
QY 5	GGGTCCGAGAGCCAGGATGGCCGCTCCGGGACCCACGCCGCGCAGGTGTCGATGCC	64		
Db 13766	GGGATCCAGCCAGCCAGGAGCAGATGCCGGCGCACCCTCGACGACACGGTGCATGCC	13825		
QY 65	GGGTCACACGGTCTCTGACGCCGCTCGTGTGATCCGTCGCG	101		
Db 13826	GCCGTCAACGGGCGCCGACGTCTGTGTGATCTCCGCTG	13862		
RESULT 5				
AAT68715				
ID	AAT68715 standard; DNA; 15872 BP.			
XX				
AC	AAT68715;			
XX				
DT	01-SEP-1997 (first entry)			
XX				
DE	Streptomyces venezuelae polyketide synthase vep ORF1.			
XX				
KM	Polyketide synthase; polyhydroxyalkanoate monomer synthase;			
KW	polyhydroxybutyrate; biodegradable polymer; vep gene;			
KW	metabolic engineering; ss.			
XX				
OS	Streptomyces venezuelae.			
XX				
EH	Key	Location/Qualifiers		
FT	CDS	20..13912		
FT		/*tag= a		
FT	CDS	14056..14136		
FT		/*tag= b		
FT	CDS	14148..15827		
FT		/*tag= c		
PN	WO9722711-A1.			
XX				
PD	26-JUN-1997.			
XX				
PF	18-DEC-1996;	96WO-US20119.		
XX				
PR	19-DEC-1995;	95US-0008847.		
XX				
PA	(MINU) UNIV MINNESOTA.			
XX				
PI	Sherman DH, Williams MD, Xue Y;			
XX				
DR	WPI; 1997-341701/31.			
DR	P-PSDB; AAM19629-30 AND AAM00918.			
XX				
PT	Expression cassettes for production of polyhydroxyalkanoate(s) -			
PT	provide wide range of biodegradable polymers for medical or			
XX	industrial use			
PS	Claim 54; Fig 23; 91pp; English.			
XX				
CC	Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide			
CC	synthase (PKS) gene cluster encoding a polyene of 12 carbons (see			
CC	also AAM19629-30 and AAM00918). It contains 5 PKS modules, with a 5'			
CC	loading module and a 3' end domain. Each of the sequenced modules			
CC	includes a keto-ACP, an acyltransferase, a dehydratase, a keto-			
CC	reductase and an acyl carrier protein domain. The gene cluster was			
CC	cloned using a heterologous hybridisation strategy from a genomic			
CC	DNA library. A novel expression cassette encoding the first module			
CC	from the vep gene cluster and module 7 from the Streptomyces ty1P			

CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC activity and can be used for PHA prodn. in host (esp. insect) cells
CC for use as a biodegradable polymer.
XX

SQ Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

Query Match Best Local Similarity 46.9%; Score 47.4; DB 18; Length 15872;

Matches 66; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 5 GCCTCCGAGCCGAGTGGCGCCCTCCG:SGCACCCACGCCGCCACGTCGATCGCC 64
DB 5150 GCGGGGAGGCGGAGTGGCGCCCTCCG:GAAAGGCTACGAGGCGATCGCCGCC 5209
OY 65 GCGGTCACGCGCTGACGCCGCTGTCGATCGTCGCGC 101
DB 5210 GCGGTCACGCGACCCACCGCGCTGTCG:CTCCGCGC 5246

RESULT 6
ID AA287283 standard; DNA; 15872 BP.

AC AA287283;

DT 05-JUN-2000 (first entry)

DE S. venezuelae vep ORF 1, SEQ ID NO:1.

KM Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM polylipier; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM chronic obstructive pulmonary disease; respiratory inflammation;
KM hypercholesterolaemia; crop protection agent; ds.

OS Streptomyces venezuelae ATCC15439.

XX Location/Qualifiers:

FT CDS

FT /tag= a

FT /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"

FT /tag= b

FT /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"

FT /tag= c

FT /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"

XX WO200000620-A2.

XX 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MINU) UNIV MINNESOTA.

XX Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI; 2000-160679/14.

XX P-PSDB; AAV77177, AAV77178, AAV77195.

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

XX synthesis of methymycin and pikromycin.

XX Example 3; Figure 23; 438bp; English.

XX The invention relates to an isolated and purified nucleic acid segment

XX comprising a desosamine biosynthetic gene cluster, a fragment or its

XX biologically active variant, where the nucleic acid sequence is not

XX derived from the eryc gene cluster of *Saccharopolyspora erythraea* or

XX streptomycetes antibiotics. The invention also relates to a macrolide

CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin, and
CC neomethymycin and narbomycin. The alternative termination or polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants. The
CC present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC sequence, designated vep ORF 1 in the specification, which actually
CC contains 3 open reading frames, which encode proteins AAV77177-77178 and
CC AAV77199. The vep ORF 1 protein is defined in the specification as a PHA
CC monomer synthase.

SQ Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

Query Match Best Local Similarity 46.9%; Score 47.4; DB 21; Length 15872;

Matches 66; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 5 GCCTCCGAGCCGAGTGGCGCCCTCCG:GACACCCACGCCGCCACGTCGATCGCC 64
DB 5150 GCGGGGAGGCGGAGTGGCGCCCTCCG:GAAAGGCTACGAGGCGATCGCCGCC 5209
OY 65 GCGGTCACGCGCTGACGCCGCTGTCGATCGTCGCGC 101
DB 5210 GCGGTCACGCGACCCACCGCGCTGTCG:CTCCGCGC 5246

RESULT 7

ID AAV21187 standard; DNA; 53789 BP.

AC AAV21187;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

KM Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

KW polyketide synthase; actinomycete; ansamycin; ds.

OS Amycolatopsis mediterranei.

XX Location/Qualifiers

FT CDS

FT /tag= a

FT /product= "polyketide synthase"

FT /tag= b

FT /product= "polyketide synthase"

FT /tag= c

FT /product= "polyketide synthase"

FT /tag= d

FT /product= "polyketide synthase"

FT /tag= e

FT /product= "polyketide synthase"

```

FT CDS 41373..51614
FT /*tag= e
FT /label= ORF_E
FT /product= "polyketide synthase"
FT 51713..5293
FT /*tag= f
FT /label= ORF_F
FT /product= "polyketide synthase"
XX
XX WO9807868-A1.
XX
XX 26-FEB-1998.
XX
XX 18-AUG-1997; 97WO-EP04495.
XX
XX 20-AUG-1996; 96EP-0810551.
XX
XX (NOVS ) NOVARTIS AG.
XX
XX Engel N, Schupp T, Toupet C;
XX
XX WPI; 1998-168172/15.
XX
XX P-PSDB; AAW52845-W52850.
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX to produce rifamycin and rifamycin analogues
XX
XX Claim 4; Page 53-102; 205pp; English.
XX
XX The present sequence represents a Amycolatopsis mediterranei rifamycin
XX synthesis gene cluster DNA fragment from the present invention. The
XX DNA fragment comprises a DNA region involved directly or indirectly
XX in the gene cluster responsible for rifamycin synthesis, including
XX the adjacent DNA regions to the right and left which, by reason of
XX their function in connection with rifamycin biosynthesis, qualify
XX as constituents of this rifamycin gene cluster, and functional
XX fragments, derivatives or constituents of these. The Amycolatopsis
XX mediterranei rifamycin synthesis gene cluster DNA fragment can be also
XX used for producing rifamycin, rifamycin analogues or precursors. It can also
XX be used for inactivating or modifying genes involved in ansamycin or
XX rifamycin biosynthesis. The DNA can be used for constructing mutant
XX actinomycetes strains from which the natural rifamycin or ansamycin
XX biosynthesis gene cluster has been partly or completely deleted. The
XX DNA fragment can be used for assembling a library of polyketide
XX synthases, which can be used for assembling a library of polyketides.
XX A hybridisation probe of the invention can be used for identifying DNA
XX fragments involved in the biosynthesis of ansamycins.
XX
XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;
XX
XX Query Match 45.1%; Score 45.6; DB 19; Length 53789;
XX Best Local Similarity 68.5%; Pred. No. 0.02;
XX Matches 63; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
XX
XX 10 CGAGGCGAGGTGGCGCCGTCGTCGACGCCGACGCGGTCGATCGCCGCGT 69
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 22698 CCAGACGACGAGTGGCTGCGCCGCTGCGCGGGGCGCGTCGATCGCCGCGT 22757
XX
XX 70 CAACGCTCCTGACGCCGTCGTGATCGCTGGCG 101
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 22758 CAACGCTCCTGACGCCGTCGTGATCGCTGGCG 22789
XX
XX RESULT 8
XX AAC55785
XX ID AAC55785 standard; DNA; 3978 BP.
XX
XX AAC55785;
XX
XX 19-JAN-2001 (first entry)
XX
XX Type I polyketide synthase orf8.
XX

```

```

KW Mitomycin: biosynthesis; mtosane ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
XX Streptomyces lavendulae.
XX
XX WO200053737-A2.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-US06394.
XX
XX 12-MAR-1999; 99US-0266965.
XX
XX (MIND ) UNIV MINNESOTA.
XX
XX (SHER/) SHERMAN D H.
XX
XX (MAOY/) MAO Y.
XX
XX (VARO/) VAROGLU M.
XX
XX (HEM/) HE M.
XX
XX (SHEL/) SHELDON P C.
XX
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
XX WPI; 2000-601980/57.
XX
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating
XX the molecular basis of mtosane ring system biosynthesis
XX
XX Example 1; Page 251-253; 399pp; English.
XX

```

```

XX This invention relates to isolated and purified nucleic acid molecules
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
XX natural products that contain a variety of functional groups, including
XX amino benzoquinone and axiridine ring systems. The S. lavendulae
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
XX spanning 55kb of DNA. The invention includes an expression cassette
XX comprising a mitomycin biosynthetic gene operably linked to a promoter,
XX and host cells transformed with the cassette. The nucleotide, and protein
XX sequences and the transformed host cells of the invention result in
XX antitastmatic, antiinflammatory, cytostatic, immunomodulatory, and
XX antibiotic activities. The nucleotide sequences are used to elucidate the
XX molecular basis for the biosynthesis of the mtosane ring system, as well
XX as to engineer the biosynthesis of novel natural products, e.g.
XX antibiotics, anti-inflammatory agents, anti-cancer agents,
XX immune-enhancers, immunosuppressants, agents to treat asthma, chronic
XX obstructive pulmonary disease as well as other disease involving
XX respiratory inflammation, or cholesterol-lowering agents or as crop
XX protection agents (e.g. fungicides or insecticides) as well as
XX biopolymers, e.g., in packaging or biomedical applications, or to engineer
XX PHA monomer synthases. Sequences AAC55782-C55811, AAC55815-C55849 and
XX AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
XX sequences and encoded proteins. Sequences AAC55812-C55814,
XX AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
XX cloning of the mitomycin biosynthetic genes.
XX
XX Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 other;
XX
XX Query Match 43.8%; Score 44.2; DB 21; Length 3978;
XX Best Local Similarity 66.0%; Pred. No. 0.05;
XX Matches 64; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
XX 5 GCGTCCGAGGCGAGGTGGCGCCCTCCGTGACACCCGACGCGCATCGATCGCC 64
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 2122 GCGACCGAGGAGGTCGCGCCCACTCGCGGACGCGCCCGCGGTGACATCGCC 2181
XX
XX 65 GCGGTCAACGCTCCTGACGCCGTCGTGATCGCTGGCG 101
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 2182 GCGGTCAACGAGCGGAGCGGTGCTGTCCGCG 2218
XX

```

RESULT 9

ID	AAC55840
XX AC	AAC55840 standard; DNA; 12249 BP.
XX XX	
XX AC	AAC55840;
XX DT	19-JAN-2001 (first entry)
DE DE	Complete Mitomycin ORF 1-9 nucleotide sequence.
KM KM	Mitomycin: biosynthesis; mitosome ring system; antibiotic; anti-cancer; anti-inflammatory; immune-enhancer; immunosuppressant; asthma; chronic obstructive pulmonary disease; respiratory inflammation; fungicide; pesticide; ds.
XX OS	Streptomyces lavendulae.
PN PN	WO200053737-A2.
PD PD	14-SEP-2000.
PF PF	10-MAR-2000; 200OWO-US06394.
PR PR	12-MAR-1999; 99US-0266965.
PA PA	(MINU) UNIV MINNESOTA. (SHER/) SHERMAN D H. (MAOY/) MAO Y. (VARO/) VAROGULU M. (HEMU/) HE M. (SHEL/) SHELDON P C.
PI PI	Sherman DH, Mao Y, Varogulu M, He M, Sheldon PC;
DR DR	WPI, 2000-601980/57.
PT PT	Novel nucleic acid molecule comprising mitomycin biosynthetic gene cluster useful for cloning mitomycin biosynthetic genes for elucidating the molecular basis of mitosome ring system biosynthesis -
PS PS	Claim 26; Figure 22; 39pp; English.
CC CC	This invention relates to isolated and purified nucleic acid molecules from the mitomycin biosynthetic gene cluster. Mitomycins are a group of natural products that contain a variety of functional groups, including amino benzocoumarone and axiridine ring systems. The <i>S. lavendulae</i> mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning 55kb of DNA. The invention includes an expression cassette comprising a mitomycin biosynthetic gene operably linked to a promoter, and host cells transformed with the cassette. The nucleotide, and protein sequences and the transformed host cells of the invention result in anthelmintic, antinflammatory, cytostatic, immunomodulatory, and antibiotic activities. The nucleotide sequences are used to elucidate the molecular basis for the biosynthesis of the mitosome ring system, as well as to engineer the biosynthesis of novel natural products, e.g., immunoenhancers, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease, or cholesterol-lowering agents or as crop protective agents (e.g. fungicides or insecticides) as well as biopolymers, e.g., in packaging or biomedical applications, or to engineer PAA monomer synthases. Sequences AAC55783-C55881, AAC55815-C55849 and AAC32485-332542 represent mitomycin biosynthetic gene cluster DNA sequences and encoded proteins. Sequences AAC55812-C55814, AAC55880-C55886 and AAC55862-C55869 represent PCR primers used in the cloning of the mitomycin biosynthetic genes.
SO SO	Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 other:
Query Match	43.8%; Score 44.2; DB 21; Length 12249;
Best Local Similarity	66.0%; Pred. No. 0.047; 33; Indels 0; Gaps 0
Matches 64;	Conservative 0; Mismatches 33; Indels 0; Gaps 0
y	5 GCCTCGAGGGCAGGATGCCGCCTTCGTGGTCAACCACAGCCCCAAGTGTCATTGCC 64

Db 3109 GCGACCGAGGACGAGGTGTGGCGCCCAACTTCGCCGACGCGCGCGCCCGGTGCACATCGCC 3168
 QY 65 GCGTCACAGCGTCTGCAGCGCTCGCAGTCGCGCGG 101
 Db 3169 GCGTCACAGGACCGGAGAGCGGTGTCTCTCTCGCGCG 3205

RESULT 10
 AAC55857
 ID AAC55857 standard; DNA; 18331 BP.
 AC
 XX AAC55857;
 DT 19-JAN-2001 (first entry)
 XX
 DE Complete nucleotide sequence of the mitomycin gene cluster.
 KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
 KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW fungicide; pesticide; ds.
 XX
 OS Streptomyces lavendulae.
 PN WO200053737-A2.
 PD 14-SEP-2000.
 XX
 PF 10-MAR-2000; 2000MO-US06394.
 XX
 PR 12-MAR-1999; 99US-0266965.
 PA (MINTU) UNIV MINNESOTA.
 PA (SHER.) SHERMAN D H.
 PA (MAOY.) MAO Y.
 PA (VARO.) VAROGLU M.
 PA (HEBM.) HE M.
 PA (SHELD.) SHELDON P C.
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
 DR WPI: 2000-601980/57.
 XX
 XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
 PT the molecular basis of mitosome ring system biosynthesis
 XX
 XX
 XX Disclosure; Figure 21: 399pp; English.
 XX
 XX This invention relates to isolated and purified nucleic acid molecules
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
 CC natural products that contain a variety of functional groups, including
 CC amino benzquinone and axiridine ring systems. The *S. lavendulae*
 CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
 CC spanning 55kb of DNA. The invention includes an expression cassette
 CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
 CC and host cells transformed with the cassette. The nucleotide, and protein
 CC sequences and the transformed host cells of the invention result in
 CC antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
 CC antibiotic activities. The nucleotide sequences are used to elucidate the
 CC molecular basis for the biosynthesis of the mitosome ring system, as well
 CC as to engineer the biosynthesis of novel natural products, e.g.
 CC antibiotics, anti-inflammatory agents, anti-cancer agents,
 CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
 CC obstructive pulmonary disease as well as other disease involving
 CC respiratory inflammation, or cholesterol-lowering agents or as crop
 CC protection agents (e.g. fungicides or insecticides) as well as
 CC biopolymers, e.g., in packaging or biomedical applications, or to engineer
 CC PAH monomer syntheses. Sequences AAC53782-C53861, AAC53815-C53849 and
 CC AAB32848-B35542 represent mitomycin biosynthetic gene cluster DNA
 CC sequences and encoded proteins. Sequences AAC53812-C53814,
 CC AAC53850-C53856 and AAC53862-C53869 represent PCR primers used in the

CC cloning of the mltomycin biosynthetic genes.
XX Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 other;
SO Query Match 43.8%; Score 44.2; DB 21; Length 18331;
Best Local Similarity 66.0%; Pred. No. 0.046;
Matches 64; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
Oy 5 GCGTCCGAGCGCCGCGCCCTCCGTCGACCCACGCCGCGACGCTGTCGATCCGC 64
Db 3109 GCGACCGGAGGAGGAGTGTCCGCACTCGCCGACGCGCCCGCGGTGACATCCGC 3168
Oy 65 GCGGTCAACGCTCTGACGCCGTCGTGATCGCTGAGCG 101
Db 3169 GCGGTCAACGAGCCGAGCGGTGTGCTGTCCGCGC 3205
RESULT 11
ID AAF24891 standard; DNA: 5544 BP.
XX AAF24891;
XX 20-APR-2001 (first entry)
XX
XX Pimaricin biosynthesis associated polyketide synthase gene.
DE
XX
XX Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW anticancer; pimaricin; ss.
XX
XX Streptomyces natalensis.
OS
XX
FH Key Location/Qualifiers
FT CDS 1..5544
FT /*tag= a
FT /product= "polyketide synthase"
XX
XX MO200077222-A1.
XX
XX 21-DEC-2000.
XX
XX 14-JUN-2000; 2000WO-EP06227.
XX
XX 14-JUN-1999; 99EP-0201893.
XX
XX (STAM) DSM NV.
XX
XX Martin JF, Aparicio JF, Collina AJ;
XX
XX WPI: 2001-080693/09.
XX P-PSDB: AAB31557.
XX
XX New polynucleotides encoding enzymes involved in the biosynthesis of
PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT the synthesis of new compounds -
XX
XX
PS Disclosure: Page 39-46; 116pp; English.
XX
XX The present sequence encodes a polyketide synthase which is associated
CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC is useful for the oxidative modification of a methyl group of a suitable
CC compound, e.g. a bioactive compound including a secondary metabolite,
CC antibiotics and anticancer agents. Recombinant cells comprising the
CC gene are useful for the production of pimaricin. The polyketide synthase
CC polynucleotide may be over expressed in Streptomyces, leading to an
CC increase in the biosynthesis of pimaricin, as a source of primers for
CC amplification reaction and as probes.
XX
XX
SQ Sequence 5544 BP; 699 A; 2052 C; 2023 G; 770 T; 0 other;
Query Match 43.0%; Score 43.4; DB 22; Length 5544;
Best Local Similarity 66.7%; Pred. No. 0.076;
Matches 62; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Oy 9 CCGAGCGCGAGTGGCGCTCCGTCGACCCACGCCGCGCGTCGATCCGCCGCG 68
Db 4037 CGGAACCCGAAATGACCCCATGCTCGCGAGTTGACGACGAGTGTGATCGCGCGC 4036
Oy 69 TCAACGTCCTGACCGCGCTGTCGATCGCTGAGCG 101
Db 4097 TCAACGTCGCGACATCCCTGTCGTGCGCGGCGC 4129
RESULT 12
ID AAQ46806 standard; DNA: 29879 BP.
XX AAQ46806;
XX
XX 22-DEC-1993 (first entry)
XX
XX eryA region of S. erythraea chromosome.
DE
XX
XX Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW erythromycin; condensation; elongation; acyl chain growth;
KW gene replacement; ss.
XX
XX Saccaropolyspora erythraea.
OS
XX
FH Key Location/Qualifiers
FT CDS 3..10199
FT /*tag= a
FT /note= "ORF 1"
FT 10218..20921
FT /*tag= b
FT /note= "ORF 2"
FT 20922..29879
FT /*tag= c
FT /note= "ORF 3"
XX
XX MO9313663-A.
XX
XX 22-JUL-1993.
XX
XX 17-JAN-1992; 92WO-US00427.
XX
XX 17-JAN-1992; 92WO-US00427.
XX
XX (ABBO) ABBOTT LAB.
XX
XX Donadio S, Katz L, McAlpine JB;
XX
XX WPI: 1993-242804/30.
XX P-PSDB: AAR44430-32.
XX
XX Biosynthesis of specific polyketide analogues esp. erythromycin
PT cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT microorganisms
XX
XX
XX Claim 27; Fig 2; 133pp; English.
XX
XX This sequence represents a fragment of the Saccaropolyspora erythraea
CC genome, designated eryA. The polypeptides encoded by this region
CC are involved in the biosynthesis of the polyketide segment of
CC erythromycin. eryA is organised in modules and each module takes care
CC of one condensation step. The precise succession of elongation steps
CC is dictated by the genetic order of the modules. This fragment may be
CC specifically altered such that novel polyketide molecules of desired
CC structure are produced. Three types of alteration may be produced:
CC those inactivating a single function in a module which does not arrest
CC acyl chain growth; those inactivating a single function in a module
CC which does affect chain growth; and those affecting an entire module.
CC The mutations may be introduced by gene replacement.
XX
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;

[illegible]

Accession	Gene	Product
AAV99230	AAV99230	standard; DNA, 1030 BP.
AAV99230	AAV99230	
09-MAR-1999	(first entry)	
DNA encoding an active acyltransferase domain (venAT) from a PKS.		
Enzymatically active acyltransferase domain; polyketide synthase; PKS; polyketide synthetase; antimicrobials; anticancer; antifungal; immunosuppressant; antineoplastic; VenAT; malonate acyltransferase domain; PKS cluster; Streptomyces venezuelae ATCC15439; ds.		
Streptomyces venezuelae.		

PS Claim 24, Fig. 13; 137bp; English:
XX
CC The present sequence encodes an enzymatically active acyltransferase
CC domain, ventP, from a polyketide synthase (PKS). The ventP protein
CC is derived from the malonate acyltransferase domain from the
CC PKS cluster of *Streptomyces venezuelae* ATCC15439. The sequence
CC can be used to replace acyltransferase domains of other PKS enzymes,
CC which are then used to generate novel polyketides by de novo biosynthesis
CC rather than by chemical modification. Polyketides of the invention have
CC the methyl groups on the macrocyclic ring substituted with -H, -Et
CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC antimicrobial, anticancer, antifungal. Polyketides of the invention are
CC antimetabolic agents, or their intermediates. immunosuppressant and/or

Sequence 1030 BP; 132 A; 383 C; 362 G; 153 T; 0 other;

	Query Match.	Score 41;	DB 20;	Length 1030;
Best Local Similarity	63.9%;	Pred. No. 0.31;		
Matches 62;	Conservative	0;	Mismatches 35;	Indels 0;
Gaps	0;			

Db 421 GCGAGCGGAGCGAGGTACGCGAGCCCTCGCGACTTCACGATGCGGTGGCC 480

07 65 GCGGTCACGGTCTCTGACGCCGTGTGTATCGCTGGCG 101

Db 481 GGAGTCAACGGCCGAAACGACCGTGTCTCGGCG 517

	RESULT 14
AA72514	
ID	AAA/72514 standard; DNA; 1030 BP.
XX	
AC	AAA/72514;
DT	20-NOV-2000 (first entry)
XX	
DE	Nucleotide sequence of the venAT domain.
XX	
KW	Erythromycin derivative; polyketide synthase; PKS; acetyltransferase; antibiotic; antifungal; anticancer; immunosuppressant; antihelminthic; venAT; ds.
XX	
OS	Streptomyces venezuelae.

PT New erythromycin derivatives produced by microorganisms with altered
 PT polyketide synthase genes -
 XX
 PS Example 9; Fig 13; 94pp; English.

This invention relates to novel erythromycin derivatives, in which the methyl groups on the macrolactone ring have been substituted with H, Et and/or OH and the ethyl side chain has been substituted with hydroxymethyl or dihydroxycyclohexylmethyl side chains. Included in the invention are vectors containing polynucleotide sequences which encode an enzymatically active acetyltransferase domain from Streptomyces. Also included are methods for altering the substrate specificity of a polypeptide synthase in a microorganism using the vectors of the invention. Polypeptides are a large class of natural products including antibiotics, antifungal, anticancer and immunosuppressant compounds. Polypeptide synthase (PRS) is a complex of enzyme activities responsible for the biosynthesis of polypeptides. The erythromycin derivatives exhibit antibiotic, antifungal, anticancer, antineoplastic and immunosuppressant activity.

The present sequence represents the nucleotide sequence encoding the ventA domain of Streptomyces venezuelae. The sequence is used in the construction of a vector of the invention.

..... SQ Sequence 1030 BP; 132 A; 383 C; 362 G; 153 T; 0 other;

Query Match	40.6%	Score 41;	DB 21;	Length 1030;
Best Local Similarity	63.9%	Pred. No. 0.31;		
Matches 62;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;

QY	5	GCGTCGAGGCGCGAGGTGGCCGCTTCGTGTGGCACCCACGCCGACGAGTGTGATCGCC	64
bB	421	GCGAGCGAGGACGAGAGTTCGCGAGGCGCCCTCGGACACTTCGACGATGCGGTTCCCTGGCC	480


```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
Query Match          96.8%; Score 97.8; DB 3; Length 68750;
Best Local Similarity 98.0%; Pred. No. 4.1e-15;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGCGGTGTCGAT 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32812 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGGTGTCGAT 32871

OY 61 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32872 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 32912
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          96.8%; Score 97.8; DB 4; Length 68750;
Best Local Similarity 98.0%; Pred. No. 4.1e-15;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGCGGTGTCGAT 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32812 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGGTGTCGAT 32871

OY 61 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32872 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 32912
```

```
RESULT 4
US-09-567-969-1
```

```
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          96.8%; Score 97.8; DB 4; Length 68750;
Best Local Similarity 98.0%; Pred. No. 4.1e-15;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGCGGTGTCGAT 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32812 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGGTGTCGAT 32871

OY 61 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32872 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 32912
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match          96.8%; Score 97.8; DB 4; Length 68750;
Best Local Similarity 98.0%; Pred. No. 4.1e-15;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGCGGTGTCGAT 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32812 CGCAGCGTCGCGAGCGCGAGTGGCGCGCTCCCTGGACACCCCAAGCGCGGTGTCGAT 32871

OY 61 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32872 CGCCGCGGTCAACGGGTCTCTGACGCGCGTGTGATTCGCTGGCG 32912
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          96.8%; Score 97.8; DB 4; Length 68750;
Best Local Similarity 98.0%; Pred. No. 4.1e-15;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCAGCGTCGAGGCCGAGGTGGCCGCTCCGTGGACACCCACGCCGCGACGGTGTGAT 60
Db 32812 CGCAGCGTCGAGGCCGAGGTGGCCGCTCCGTGGACACCCACGCCGCGACGGTGTGAT 32871
QY 61 CGCCGCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 101
Db 32872 CGCCGCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 32912

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          96.8%; Score 97.8; DB 4; Length 68750;
Best Local Similarity 98.0%; Pred. No. 4.1e-15;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCAGCGTCGAGGCCGAGGTGGCCGCTCCGTGGACACCCACGCCGCGCGACGGTGTGAT 60
Db 32812 CGCAGCGTCGAGGCCGAGGTGGCCGCTCCGTGGACACCCACGCCGCGCGACGGTGTGAT 32871
QY 61 CGCCGCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 101
```

```
Db 32872 CGCCGCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 32912

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          96.8%; Score 97.8; DB 4; Length 68750;
Best Local Similarity 98.0%; Pred. No. 4.1e-15;
Matches 99; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCAGCGTCGAGGCCGAGGTGGCCGCTCCGTGGACACCCACGCCGCGCGACGGTGTGAT 60
Db 32812 CGCAGCGTCGAGGCCGAGGTGGCCGCTCCGTGGACACCCACGCCGCGCGACGGTGTGAT 32871
QY 61 CGCCGCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 101
Db 32872 CGCCGCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 32912

RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match          46.3%; Score 47.4; DB 4; Length 15872;
Best Local Similarity 68.0%; Pred. No. 0.0019;
Matches 66; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 5 GCGTCGAGCGCCGAGGTGGCCGCTCCGTGGACACCCACGCCGCGCGACGGTGTGATGCC 64
Db 5150 GCGGCGAGGCGGAGGTGGCCGCTCCGTGGACACCCACGCCGCGCGACGGTGTGATGCC 5209
QY 65 GCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 101
Db 32872 CGCCGCGGTCAACGGTCTGTGACGCCGCTGTGATCGTGGCG 32912
```

DB 5210 GCCGTCACGACGACCGCGCTGTC TCCTCCGGCG 5246

RESULT 10

US-07-642-734C-1

; Sequence 1, Application US/07642734C;

; Patent No. 5824513

; GENERAL INFORMATION:

; APPLICANT: Katz, L

; APPLICANT: Donadio, S

; APPLICANT: McAlpine, J B

; TITLE OF INVENTION: Recombinant DNA Method for Producing

; TITLE OF INVENTION: Erythromycin Analogs

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Edward H. Gorman

; STREET: Abbott Laboratories D377/AP6D-2 One Abbott

; STREET: Park Rd

; CITY: Abbott Park

; STATE: IL

; COUNTRY: US

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/642,734C

; FILING DATE: 17-JAN-91

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Danckers, Andreas M

; REGISTRATION NUMBER: 32652

; REFERENCE/DOCKET NUMBER: 4952.U;:01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 708-937-9396

; TELEFAX: 708-938-2623

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11219 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Saccharopolyspora erythraea

; STRAIN: NRRL 2338

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 744..6559

; OTHER INFORMATION: /function= "APPROXIMATE SPAN OF

; OTHER INFORMATION: MODULE 1"

; OTHER INFORMATION: /label= FUNCTION

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 744..11219

; OTHER INFORMATION: /function= "gene= "eryA"

; OTHER INFORMATION: /product= "Ori encoding modules 1 & 2 for

; OTHER INFORMATION: 6-deoxyerythronolide B"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 744..1868

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: acyltransferase" domain 1 of module 1"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 1998..2198

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: acyl carrier domain 1 of module 1"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 2250..3626

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 3831..4811

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: acyltransferase 2 domain of module 1"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 5574..6125

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: beta-ketoreductase domain of module 1"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 6369..6626

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: acyl carrier domain 2 of module 1"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 6678..11219

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: module 2"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 6678..8066

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 8262..9305

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: acyltransferase domain of module 2"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 9906..10454

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: beta-ketoreductase domain of module 2"

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: 10707..10964

; OTHER INFORMATION: /function= "approximate span of

; OTHER INFORMATION: acyl carrier domain of module 2"

; US-07-642-734C-1

Query Match 40.8%; Score 41.2; DB 1; Length 11219;

Best Local Similarity 64.9%; Pred. No. 0.053; 33; Indels 0; Gaps 0;

Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 8 TCCGAGCGCGAGCTGCGCCGCTGGCGACCCGCGCGCGCTGCGATCGCGCG 67

DB 8673 TCCGAGCTGCGAGAGCTGCTGCCGACGACGACTGAGCGCGAGGCTGCCGCG 8732

QY 68 GTCAAGCTCTGACGCGCGCTGATCGCTGCG 101

DB 8733 GTCAAGCGCCCGACGCGCTGCTGCGCGGTG 8766

RESULT 11

US-08-439-009A-1

; Sequence 1, Application US/08439009A

; Patent No. 6004787

; GENERAL INFORMATION:

; APPLICANT: Donadio, S

; APPLICANT: Katz, L

; APPLICANT: McAlpine, J B

; TITLE OF INVENTION: Method of Directing Biosynthesis of

; TITLE OF INVENTION: Specific Polyketides

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Steven F. Weinstock

; STREET: Abbott Laboratories D377/AP6D-2 One Abbott

STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..11219
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: MODULE 1"
OTHER INFORMATION: /label=FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function="gene="eryA"
OTHER INFORMATION: /product="ORF1 encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..1868
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2198
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2250..3626
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3831..4811
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5574..6125
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6369..6626
OTHER INFORMATION: /function="approximate span of

OTHER INFORMATION: acyl carrier domain 2 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..11219
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..8066
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8262..9305
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9906..10454
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10707..10964
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 2"
US-08-439-009A-1
Query Match 40.8%; Score 41.2; DB 3; Length 11219;
Best Local Similarity 64.9%; Pred. No. 0.053;
Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 8 TCCGAGGCCGAGGTGCCGCTCCGTGGACACCCACGCCGCCGAGGTGTGATGCCGCCG 67
DB 8673 TCCGACGTCGAGAGCTGCTGCCGACACGACGTCGACGCGAGCGTGGAGTGCGCCG 8732
QY 68 GTCAACGCTCCTGACGCGCGTGTGATGCTGCG 101
DB 8733 GTCAACGCGCCCGACGCGTGTGTGTCGCCGCTG 8766
RESULT 12
US-08-858-003-2
Sequence 2, Application US/08858003
Patent No. 6060234
GENERAL INFORMATION:
APPLICANT: Katz, Leonard
APPLICANT: Staszi, Diane L.
APPLICANT: Summers Jr., Richard G.
APPLICANT: Ruan, Xiaolan
APPLICANT: Pereda-Lopez, Ana
TITLE OF INVENTION: NOVEL POLYPEPTIDE DERIVATIVES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd.
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,003
FILING DATE: 16-MAY-1979
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-858-003-2

Query Match 40.6%; Score 41; DB 3; Length 1030;
Best Local Similarity 63.9%; Pred. No. 0.067;
Matches 62; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 5 GCGTCCGAGCCGAGGTGGCCGCTCGCTGACACCCGACGCCGCGTGTGATGCGC 64
DB 421 GCGAGCGAGGAGGAGGTGACGAGGCGCCCTCCGACCTTCGATGCGTGGCGC 480
QY 65 GCGGTCAACGGCTCTGACGCCGCTGATGCTGCGCG 101
DB 481 GGAATCAACGGCCGAGACGACCGCTGCTCGCGCG 517

RESULT 13
US-09-078-166-2
Sequence 2, Application US/09078166
Patent No. 6063561
GENERAL INFORMATION:
APPLICANT: Katz, Leonard
APPLICANT: Stassell, Diane L.
APPLICANT: Summers Jr., Richard G.
APPLICANT: Ruan, Xiaojun
APPLICANT: Pereda-Lopez, Ana
APPLICANT: Kakavas, Stephen J.
TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd.
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/078,166
FILING DATE: 16-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-078-166-2

Query Match 40.6%; Score 41; DB 3; Length 1030;
Best Local Similarity 63.9%; Pred. No. 0.067;
Matches 62; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 5 GCGTCCGAGCCGAGGTGGCCGCTCGCTGACACCCGACGCCGCGTGTGATGCGC 64
DB 421 GCGAGCGAGGAGGAGGTGACGAGGCGCCCTCCGACCTTCGATGCGTGGCGC 480
QY 65 GCGGTCAACGGCTCTGACGCCGCTGATGCTGCGCG 101
DB 481 GGAATCAACGGCCGAGACGACCGCTGCTCGCGCG 517

RESULT 14
US-08-997-467-2
Sequence 2, Application US/08997467
Patent No. 620813
GENERAL INFORMATION:
APPLICANT: Katz, Leonard
APPLICANT: Stassell, Diane L.
APPLICANT: Summers Jr., Richard G.
APPLICANT: Ruan, Xiaojun
APPLICANT: Pereda-Lopez, Ana
APPLICANT: Kakavas, Stephen J.
TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Rd.
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,467
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/858,003
FILING DATE: 16-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuto
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 4952.US.P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847)-938-3137
TELEFAX: (847)-938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-997-467-2

Query Match 40.6%; Score 41; DB 4; Length 1030;
Best Local Similarity 63.9%; Pred. No. 0.067;
Matches 62; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 5 GCGTCCGAGCCGAGGTGGCCGCTCGCTGACACCCGACGCCGCGTGTGATGCGC 64

1
2
3
4

=

—


```

; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR APPLICATION NUMBER: 60/269,306
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4036
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-4036

Query Match
Best Local Similarity 33.5%; Score 33.8; DB 10; Length 546;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 CGCAGCGTCGAGCGCGAGGTGGCGGCTTCGTGGCACCCCGCCGACGGTGTGATC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 CGCTTCGACCCCTGCGCGAGTGGCGGCTTCGTGGACCGCCCGCGCGAT 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 CGCCGCGTCACGCTCTGACGCCGCTGCTGATCGTGGCG 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 CGCCGCGCGCGCCGACGACGAGGATGCGCGCTACG 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
; US-09-850-964-4
; Sequence 4, Application US/09850964
; Patent No. US20020104121A1
; GENERAL INFORMATION:
; APPLICANT: MCELROY, DAVID
; APPLICANT: KRIZ, ALAN L.
; APPLICANT: OROZCO, EMIL
; APPLICANT: GRIFFOR, MART
; TITLE OF INVENTION: THE MAIZE A3 PROMOTER AND RICE ACTIN 2 INTRON AND
; FILE REFERENCE: DEKM.166
; CURRENT APPLICATION NUMBER: US/09/850,964
; PRIOR FILING DATE: 2001-05-07
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1294
; TYPE: DNA
; ORGANISM: Zea mays/rice
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (883)..(972)
; FILE REFERENCE: ELITRA.011A
; OTHER INFORMATION: N - A or C or G or T
```

```

US-09-850-964-4
Query Match
Best Local Similarity 32.5%; Score 32.8; DB 10; Length 1294;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 2 GCAGCGTCGAGCGCGAGGTGGCGGCTTCGTGGCACCCCGCCGACGGTGTGATC 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 GCGGTAGCGCGCGCGACGGTGGCGGCTTCGTGGCACCCCGCCGACGGTGTGATC 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 62 GCGCGGTCAAGCGCTCTGACGCCGCTGCTGATCGTGGCG 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 GCGCGCACCGTAGCGCGCGCGCGCGCGCGCGCGTAGCGCGCG 638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
; US-09-850-964-3
; Sequence 3, Application US/09850964
; Patent No. US20020104121A1
; GENERAL INFORMATION:
; APPLICANT: MCELROY, DAVID
; APPLICANT: KRIZ, ALAN L.
; APPLICANT: OROZCO, EMIL
; APPLICANT: GRIFFOR, MART
; TITLE OF INVENTION: THE MAIZE A3 PROMOTER AND RICE ACTIN 2 INTRON AND
; FILE REFERENCE: DEKM.166
; CURRENT APPLICATION NUMBER: US/09/850,964
; PRIOR FILING DATE: 2001-05-07
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2289
; TYPE: DNA
; ORGANISM: Zea mays/rice
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (889)..(1825)
; OTHER INFORMATION: N - A or C or G or T
; US-09-850-964-3

Query Match
Best Local Similarity 32.5%; Score 32.8; DB 10; Length 2289;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 2 GCAGCGTCGAGCGCGAGGTGGCGGCTTCGTGGCACCCCGCCGACGGTGTGATC 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GCGGTAGCGCGCGCGACGGTGGCGGCTTCGTGGCACCCCGCCGACGGTGTGATC 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 62 GCGCGGTCAAGCGCTCTGACGCCGCTGCTGATCGTGGCG 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 GCGCGCACCGTAGCGCGCGCGCGCGCGCGCGCGTAGCGCGCG 644
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
; US-09-815-242-7987
; Sequence 7987, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
```

```

CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7987
LENGTH: 1530
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1530)
OS-09-815-242-7987

```

Query Match	31.9%	Score 32.2	DB 10	Length 1530
Best Local Similarity	66.7%	Pred. No. 1.8		
Matches 46	Conservative 0	Mismatches 23	Indels 0	Gaps 0

OY 5 GCGTCCGAGGCCGAGGTGGCGGCCTTCGTGGCACCCCACGCCGCAACGGTGTCGATCGCC 64
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 952 GCGCGCAGAGGCGACGTGATCTCCGGCGGCACCTTCACAGCGGAACGGTGCGCATGGCC 1011

QY	65	GGGTCAC	73
Db	1012	GGCGACAAC	1020

```

RESULT 6
US-09-815-242-7769
: Sequence 7769, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsem, Kari L.
: APPLICANT: Zykkind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: E11TRA.011A
: CURRENT APPLICATION NUMBER: US/09/815.242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0

```

```

; SEQ ID NO 7769
;
; LENGTH: 1395
;
; TYPE: DNA
;
; ORGANISM: Pseudomonas aeruginosa
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: (1)..(1395)
;
US-09-815-242-7769

```

Query Match	31.7%	Score 32	DB 10	Length 1395
Best Local Similarity	60.2%	Pred. No. 2.1		
Matches 53	Conservative 0	Mismatches 35	Indels 0	Gaps 0

Qy 14 GCCGAGGTGGCCGCCCTCCGTGGCACCACCGCGGCACGGTGTGATGCGCGCGGTCAAC 73
||||| |||| | |||| | |||| | |||| | |||| | |||| |
Db 988 GCGGAGTGATTGCGCGGCAGAAGCTGCCAGTTGCCCCGGTGGCGATCCC CGGGGTGTC 1047

OY	74	GGTCCTGACGCCGTCGTGATCGCTGGCG	101
Db	1048	TTCAACCGATCCGGAGTGGTGGCGCCG	1075

RESULT 7
US-09-861-289-36

; Sequence 36, Application US/09861289
; Patent No. US20020110897A1

```

; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```

APPLICANT: Liu, H.
APPLICANT: Xue, Y.

```

; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILING NUMBER: 600 428051

```

```

; FILE REFERENCE: 600.4380S1
; CURRENT APPLICATION NUMBER: US/09/861,285
; CURRENT FILING DATE: 2001-05-18

```

;; CURRENT FILING DATE: 2001-03-18
;; PRIOR APPLICATION NUMBER: 09/105,537
;; PRIOR FILING DATE: 1998-06-26

```

; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows

```

```

; SOFTWARE. EAS
; SEQ ID NO 36
; LENGTH: 4041

```

/ ORGANISM: Streptomyces venezuelae
 ; TYPE: DNA
 ;

US-09-861-289-36

Query Match 31.78; Score 32; DB 10; Length 4041;

Best Local Similarity 79.2%; Pred. No. 2;
Matches 38; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

[illegible]

RESULT 8
US-09-86

; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; CHUNYI TECHNOLOGY

; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; ADDITIONAL INFORMATION:

APPLICANT: Liu, Y
APPLICANT: Xue, Y
APPLICANT: Zhao

; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600 438us1

```

; FILE REFERENCE: 000.438051
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18

```

; COMMENT FILING DATE: 2001-03-10
 ; PRIOR APPLICATION NUMBER: 09/105,537
 ; PRIOR FILING DATE: 1998-06-26

```

; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0

```

```

; SEQ ID NO 5
; LENGTH: 36778

```

```

; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
us-09-861-289-5

```

Query Match	31.7%	Score 32;	DB 10;	Length 36778;
Best Local Similarity	79.2%	Prel. NO. 1.8;		
Matches 38; Conservative	0;	Mismatches 10;	Indels 0;	Gaps 0;

```

QY      54  TGTGATCGCCCGGGTCAACGTCCTG/CGCCGTCGTGATCGCTGGCC 101
          |||||
Db 33913 TGTGATCGCCCGGGTCAACGCGGCTAC/CGCCACCGTGTGTTTCGGGCG 33960
          |||||

```

RESULT 9
US-10-044-090-309
; Sequence 309, Application US/10044093
; Patent No. US20020137081A1

```

1  APPLICANT: Olga Bandman
2  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
3  FILE REFERENCE: PA-0028 US
4  CURRENT APPLICATION NUMBER: US/10/04,090
5  CURRENT FILING DATE: 2002-01-09
6  NUMBER OF SEQ. ID NOS: 850
7  SOFTWARE: PERL Program
8  SEQ. ID NO 309

```

```

? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: Incyte ID No. US20020137081A1 300437.18
? NAME/KEY: unsure
? LOCATION: 2060
? OTHER INFORMATION: a, t, c, g, or o:her
? US-10-044-090-309

```

Query Match	31.5%;	Score 31.8;	DB 12;	Length 2724;
Best Local Similarity	67.28;	Pred. NO. 2.3;		
Matches 45;	Conservative 0;	Mismatches 22;	Indels 0;	Gaps 0;

Oy 3 CAGCGTCCGAGGCCGAGGTGGCCGCCTCC(ATTGGACCCCCCAGCCGCCGAGGTTCGATCC 62
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 601 CCGCGCCGAGGCGCGCGCCCGCCTCC(GCCGCACTATCGCCGCGCGCGGCCACCG 660

ay	63	CCGCGGT	69
Db	661	CCGTCGT	667

RESULT 10
 US-09-821-167-1/c
 ; Sequence 1, Application US/09821167
 Patent No. US20020015989A1
 GENERAL INFORMATION:
 APPLICANT: Hosted Jr., Thomas J.
 APPLICANT: Horan, Ann C.
 TITLE OF INVENTION: Isolated Nucleic Acids from *Micromonospora rosalia*
 TITLE OF INVENTION: plasmid pMR2 and Vectors Made Therefrom
 FILE REFERENCE: IN011490
 CURRENT APPLICATION NUMBER: US/09/821,167
 CURRENT FILING DATE: 2001-03-29
 PRIOR APPLICATION NUMBER: US 60/194,461
 PRIOR FILING DATE: 2000-04-04
 NUMBER OF SEQ ID NOS: 15
 SOFTWARE: PatentIn Ver. 2.1

```

: .LENGTH: 11188
: TYPE: DNA
: ORGANISM: Micromonospora rosalia
: FEATURE:
: NAME/KEY: RBS
: LOCATION: (6055)..(6059)

```

```

? NAME/KEY: RBS
? LOCATION: (6391)..(6394)
? NAME/KEY: RBS
? LOCATION: (8084)..(8088)
? NAME/KEY: RBS
? LOCATION: (9834)..(9837)
? NAME/KEY: RBS
? LOCATION: (10010)..(10012)
?
US-09-821-167-1

```

Query Match	31.5%	Score 31.8;	DB 10;	Length 1118;
Best Local Similarity	59.3%	Pred. NO. 2.1;		
Matches	54;	Conservative	0;	Mismatches 37;
			Indels	0;
			Gaps	0;

QY 11 GAGGCCGAGGTGGCCGCTCTCGTGGCACCCACGCGCCACACGGTGTGATCGCGCGGTC 70
||||||| - - - ||| - - - ||| ||||| - - - - -
Db 1507 GAGGCCGAGCGGGGTGAAGCGCGCCAGAGGACGCGGCACACGGTGTGCTTTTGAAGCTG 1448

```

QY      71 AACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          | | | | | | | | | | | | | | | |
Db     1447 AGCGTGTGACGACATCGGGCTATCTCGCG 1417

```

RESULT 11
US-09-880-107-2102/c
; Sequence 2102, Application US/09880107
; Patent No. US20020142981A1

```

1  APPLICANT : Horne, Darci T.
2  APPLICANT : Vockley, Joseph G.
3  APPLICANT : Scherf, Uwe
4  APPLICANT : Gene Logic, Inc.
5  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
6  FILE REFERENCE: 44921-5028-MO
7  CURRENT APPLICATION NUMBER: US/09/880,107
8  PRIOR FILING DATE: 2001-06-14
9  PRIOR APPLICATION NUMBER: US 60/211,379
10 PRIOR FILING DATE: 2000-06-14
11 PRIOR APPLICATION NUMBER: US 60/237,054
12 PRIOR FILING DATE: 2000-10-02
13 NUMBER OF SEQ ID NOS: 3950
14 SOFTWARE: PatentIn Ver. 2.1
15 SEQ ID NO 2102
16 LENGTH: 2038
17 TYPE: DNA
18 ORGANISM: Homo sapiens
19 FEATURE:
20 OTHER INFORMATION: Genbank Accession No. US20020142981A1 J04029
21 US-09-880-107-2102

```

Query Match	30.98;	Score 31.2;	DB 10;	Length 2038;
Best Local Similarity	57.08;	Pred. No. 3.2;		
Matches	57;	Conservative	0;	Mismatches 43;
			Indels	0;
			Gaps	0;

Qy 2 GCACGCTCCGAGGCGCGAGTGGCCCCCTCCGTGGACCCCAACGCCACGGTGCATC 61
||| ||| ||| | ||||| ||| ||| ||||| ||| |
Db 1561 GCTCTGCCGCGCGCGAGCTGCCGCCCCCGTAGCGCGCGCGCGCGCGAACTGCC 1502

QY 62 GCCCGGTCACCGGTCCTGACGCCGTCGTATCGCTGGCG 101
 ||||| ||||| ||||| |||||
 Db 1501 ACCACCGTAGCCGCGCTGGAAC TGCCGCCGCTGGCCGCG 1462

RESULT 12
US-09-861-289-34/c
; Sequence 34, Application US/09861289

1. APPLICANT: Sherman, D. H.
 2. APPLICANT: Liu, H.
 3. APPLICANT: Xue, Y.
 4. APPLICANT: Zhao, L.
 5. TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.438051
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 4689
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-34

Query Match
Best Local Similarity 30.9%; Score 31.2; DB 10; Length 4689;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 2 GCAGGCTCCGAGAGCCGAGGTGCGCCCTCCGTGCGACCCACGCCGCGGTGTCATC 61
DB 3444 GCGGCTGACGAGAGAGAGGTGTCGCTCCGCTGCGGCATCCAGCTGCGCGTGGCT 3385
OY 62 GCCGCGGTCAACGCTCTGACGCCCTCGTATCGCTGGCG 101
DB 3384 GCCGAGGGCTCCGCTGCGCCGCGGTGATGAGACGGTGGCG 3345

RESULT 13
US-09-880-107-3699/c
Sequence 3699, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Scherf, Joseph G.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-MO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3699
LENGTH: 6483
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X14487
US-09-880-107-3699

Query Match
Best Local Similarity 30.9%; Score 31.2; DB 10; Length 6483;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 2 GCAGGCTCCGAGAGCCGAGGTGCGCCCTCCGTGCGACCCACGCCGCGGTGTCATC 61
DB 5407 GCTGCTGCGCGCGGAGAGAGGTGTCGCTCCGCTGCGGCATCCAGCTGCGCGTGGCT 5348
OY 62 GCCGCGGTCAACGCTCTGACGCCCTCGTATCGCTGGCG 101
DB 5347 ACCACCGTAGCCGCGCTGGAATCTCCGCGGTGGCGCGG 5308

RESULT 14
US-09-861-289-30/c
Sequence 30, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.

APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438051
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match
Best Local Similarity 30.9%; Score 31.2; DB 10; Length 13842;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 2 GCAGGCTCCGAGAGCCGAGGTGCGCCCTCCGTGCGACCCACGCCGCGGTGTCATC 61
DB 6720 GCGGCTGACGAGAGAGAGGTGTCGCTCCGCTGCGGCATCCAGCTGCGCGTGGCT 6661
OY 62 GCCGCGGTCAACGCTCTGACGCCCTCGTATCGCTGGCG 101
DB 6660 GCCGAGGGCTCCGCTGCGCCGCGGTGATGAGACGGTGGCG 6621

RESULT 15
US-09-861-289-5/c
Sequence 5, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438051
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 36778
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match
Best Local Similarity 30.9%; Score 31.2; DB 10; Length 36778;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 2 GCAGGCTCCGAGAGCCGAGGTGCGCCCTCCGTGCGACCCACGCCGCGGTGTCATC 61
DB 30434 GCGGCTGACGAGAGAGAGGTGTCGCTCCGCTGCGGCATCCAGCTGCGCGTGGCT 30375
OY 62 GCCGCGGTCAACGCTCTGACGCCCTCGTATCGCTGGCG 101
DB 30374 GCCGAGGGCTCCGCTGCGCCGCGGTGATGAGACGGTGGCG 30335

Search completed: November 5, 2002, 23:01:40
Job time : 104.254 secs


```
/clone="PM047f04.r"
/clone.lib="Porphyra yezoensis TU-1"
/note="Vector: phluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      64 a      209 c      153 g      81 t
ORIGIN
Query Match      41.2%; Score 41.6; DB 10; Length 507;
Best Local Similarity 67.0%; Pred. No. 4.2; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 29;
OY 1 CGCAGCGTCCGAGCGGAGGTGCGCCCTCGTGGACCCACGCCGCGCAGGTGCGAT 60
Db 229 CGCGCGGGGGGTGCCATGCTGTCGCCCTGCGGCCGCCACCGCGCGCGCGCCG 288
OY 61 CGCGCGGTCAACAGGTCTGACGCCGTC 88
Db 289 CACCGCAGTCCGCGCTCTGTCGCCGC 316

RESULT 2
AV434343      512 bp      mRNA      linear      EST 23-AUG-2000
LOCUS      AV434343 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION      PM041h03_r 5', mRNA sequence.
ACCESSION      AV434343
VERSION      AV434343.1 GI:8589568
KEYWORDS      EST.
SOURCE      Porphyra yezoensis.
ORGANISM      Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
REFERENCE      1 (bases 1 to 512)
AUTHORS      Nikaïdo,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
Tabata,S.
TITLE      Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL      DNA Res. 7, 223-227 (2000)
MEDLINE      20363100
COMMENT      Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Source
1..512
/organism="Porphyra yezoensis"
/strain="TU-1"
/db.xref="taxon:2788"
/clone.lib="PM041h03.r"
/note="Vector: phluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      66 a      209 c      153 g      84 t
ORIGIN
Query Match      41.2%; Score 41.6; DB 10; Length 512;
Best Local Similarity 67.0%; Pred. No. 4.2; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
OY 1 CGCAGCGTCCGAGCGGAGGTGCGCCCTCGTGGACCCACGCCGCGCAGGTGCGAT 60
Db 229 CGCGCGGGGGGTGCCATGCTGTCGCCCTGCGGCCGCCACCGCGCGCGCGCCG 290
OY 61 CGCGCGGTCAACAGGTCTGACGCCGTC 88
Db 291 CACCGCAGTCCGCGCTCTGTCGCCGC 316

RESULT 3
AV431953      520 bp      mRNA      linear      EST 23-AUG-2000
LOCUS      AV431953 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION
```

```
PM008a11_r 5', mRNA sequence.
ACCESSION      AV431953
VERSION      AV431953.1 GI:8587178
KEYWORDS      EST.
SOURCE      Porphyra yezoensis.
ORGANISM      Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
REFERENCE      1 (bases 1 to 520)
AUTHORS      Nikaïdo,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
Tabata,S.
TITLE      Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL      DNA Res. 7, 223-227 (2000)
MEDLINE      20363100
COMMENT      Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Source
1..520
/organism="Porphyra yezoensis"
/strain="TU-1"
/db.xref="taxon:2788"
/clone.lib="PM008a11.r"
/note="Vector: phluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      66 a      213 c      157 g      84 t
ORIGIN
Query Match      41.2%; Score 41.6; DB 10; Length 520;
Best Local Similarity 67.0%; Pred. No. 4.2; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
OY 1 CGCAGCGTCCGAGCGGAGGTGCGCCCTCGTGGACCCACGCCGCGCAGGTGCGAT 60
Db 235 CGCGCGGGGGGTGCCATGCTGTCGCCCTGCGGCCGCCACCGCGCGCGCGCCG 294
OY 61 CGCGCGGTCAACAGGTCTGACGCCGTC 88
Db 295 CACCGCAGTCCGCGCTCTGTCGCCGC 322

RESULT 4
AV431696      528 bp      mRNA      linear      EST 23-AUG-2000
LOCUS      AV431696 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION      PM004c09_r 5', mRNA sequence.
ACCESSION      AV431696
VERSION      AV431696.1 GI:8586921
KEYWORDS      EST.
SOURCE      Porphyra yezoensis.
ORGANISM      Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
REFERENCE      1 (bases 1 to 528)
AUTHORS      Nikaïdo,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
Tabata,S.
TITLE      Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL      DNA Res. 7, 223-227 (2000)
MEDLINE      20363100
COMMENT      Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
Source
1..528
/organism="Porphyra yezoensis"
/strain="TU-1"
```

```

/db_xref="taxon:2788"
/clone="PM004c09_r"
/clone_lib="Porphyra yezoensis TU-1"
/Note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      71 a      211 c      160 g      86 t
ORIGIN
Query Match      41.2%; Score 41.6; DB 10; Length 528;
Best Local Similarity 67.0%; Pred. No. 4.2;
Matches 59; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 CGCAGCGTCCAGGCGGAGGCGGCTCGTGGCACCACCGCGCAGGTGCGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 CGCGCGCGGCGTGGCAGCTGCTGTCGCGTGGCCGCCACCGCGCAGCGCGCC 265
QY 61 CGCCGCGGTCAACGCTCTGACGCCGTC 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 CACCGCAGTCCGCGCTCTGCTGCCGCC 293

RESULT 5
AV431496
LOCUS      AV431496      530 bp      mRNA      linear      EST 23-AUG-2000
DEFINITION      AV431496 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
PM001c08_r 5', mRNA sequence.
ACCESSION      AV431496
VERSION      AV431496.1 GI:8586721
KEYWORDS      EST.
SOURCE      Porphyra yezoensis.
ORGANISM      Porphyra yezoensis
              Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
              Porphyra.
REFERENCE      1 (bases 1 to 530)
AUTHORS      Nkaido,T., Asamizu,E., Nakajima,M., Nakamura,Y., Sage,N. and
              Tabata,S.
TITLE      Generation of 10,154 expressed sequence tags from a leafy
              gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL      DNA Res. 7, 223-227 (2000)
MEDLINE      20363100
COMMENT      Contact: Erika Asamizu
              The First Laboratory for Plant Gene Research
              Kazusa DNA Research Institute
              Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
              Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
    source
        1..530
            /organism="Porphyra yezoensis"
            /strain="TU-1"
            /db_xref="taxon:2788"
            /clone="PM001c08_r"
            /clone_lib="Porphyra yezoensis TU-1"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT      67 a      214 c      167 g      82 t
ORIGIN
Query Match      41.2%; Score 41.6; DB 10; Length 530;
Best Local Similarity 67.0%; Pred. No. 4.2;
Matches 59; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 CGCAGCGTCCAGGCGGAGGCGGCTCGTGGCACCACCGCGCAGGTGCGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 CGCGCGCGGCGTGGCAGCTGCTGTCGCGTGGCCGCCACCGCGCAGCGCGCC 249
QY 61 CGCCGCGGTCAACGCTCTGACGCCGTC 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 CACCGCAGTCCGCGCTCTGCTGCCGCC 277

RESULT 6
BF630328/c
LOCUS      BF630328      495 bp      mRNA      linear      EST 22-OCT-2001

```

```

DEFINITION      HVSMEB0008M06f Hordeum vulgare seedling shoot EST library
HVCNDA0002 (Dehydration stress) Hordeum vulgare cDNA clone
HVSMEB0008M06f, mRNA sequence.
ACCESSION      BF630328
VERSION      BF630328.2 GI:13090843
KEYWORDS      EST.
SOURCE      Hordeum vulgare.
ORGANISM      Hordeum vulgare
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
              ; Triticeae; Hordeum.
REFERENCE      1 (bases 1 to 495)
AUTHORS      Wing,R., Close,T.J., Kleinof,A., Wise,R., Begum,D., Frisch,D., Yu
              ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
              ,R.D., Oates,R. and Main,D.
TITLE      Development of a genetically and physically anchored EST resource
              for barley genomics: Morex drought-stressed seedling shoot cDNA
              library
JOURNAL      Unpublished (2001)
COMMENT      On Dec 19, 2000 this sequence version replaced gi:11894486.
              Contact: Wing RA
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 7288
              Fax: 864 656 4293
              Email: twing@clemson.edu
              Total hg bases = 395
              Seq primer: AATTACCTTCACCTAAGCG
              High quality sequence stop: 484.
              Location/Qualifiers
                  1..495
                      /organism="Hordeum vulgare"
                      /cultivar="Morex"
                      /db_xref="taxon:4513"
                      /clone="HVSMEB0008M06f"
                      /clone_lib="Hordeum vulgare seedling shoot EST library
                      HVCNDA0002 (Dehydration stress)"
                      /tissue="type="Seedling shoot"
                      /lab="host="TUC121"
                      /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
                      Seeds were surface sterilized then germinated under axenic
                      conditions in the dark at room temperature on filter paper
                      with water, nystatin and cefotaxime in covered
                      crystallization dishes. Five-day old seedlings were
                      incubated at 90° RH for 24 hr. Shoots were then harvested,
                      total RNA was prepared, poly(A) RNA was purified, one
                      primary unamplified cDNA library was made, 600000 pfu were
                      in vivo excised to give pBluescript SK(-) cDNA phagemids.
                      These steps were performed in the TU Close laboratory at
                      the University of California, Riverside (Choi, Close,
                      Fenton). Phagemids were plated and picked at the Clemson
                      University Genomics Institute (CUGI) (Begum, Palmer,
                      Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                      sequencing and sequence analysis were performed at CUGI
                      (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                      The sequence has been trimmed to remove vector sequence
                      and contains a minimum of 100 bases of phred value 20 or
                      above. For more details on library preparation and
                      sequence analysis see
                      http://www.genome.clemson.edu/projects/barley. To order
                      this clone see http://www.genome.clemson.edu/orders Also
                      see Close TJ, Wing R, Kleinof A, Wise R (2001)
                      Genetically and physically anchored EST resources for
                      barley genomics. Barley Genetics Newsletter 31:29-30.
                      (http://wheat.pw.usda.gov/gpapes/bgn/31/cover.html)"
BASE COUNT      78 a      146 c      198 g      73 t
ORIGIN
Query Match      40.6%; Score 41; DB 12; Length 495;
Best Local Similarity 63.9%; Pred. No. 5.7;
Matches 62; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```


Best Local Similarity	62.48;	Pred. No.	8.6;
Matches	63;	Conservative	0;
		Mismatches	38;
		Indels	0;
		Gaps	0;

Dy 1 CGCAGCGTCCGAGGCCGAGTGCCGCCTCCGTGGCACCCACGCGGCCACGCTGCAT 60
| | | | | | | | | | | | | | | | | |
Db 230 CACCGGTCCTGGCGCGCGCGTAGCCACTCCCTGGCGCGCGCTAGCGGCCACCGGCTGT 171

QY	61	CGCCCGGGTAAACGGTCTCTGACGCCGTCGTGATGCTGGCG	101
Db	170	AGCCTCCGCCACACGGCGCGCTGCGCGCGCTAGCGGCCGCG	130

RESULT 9
BE215355/c

LOCUS	883 bp	mRNA	linear	EST 23-OCT-2001
BE215355				
HY_CBD0006J05f		Hordeum vulgare seedling	green leaf	EST library
HYCDNA0005		(Blumeria challenged)	Hordeum vulgare	cDNA clone
HY_CBD0006J05f		mRNA sequence.		

ACCESSION	BE215355
VERSION	BE215355.2
KEYWORDS	GI:13264386 EST.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.

REFERENCE	1 (bases 1 to 883)
AUTHORS	Wing, R., Close, T.J., Kleinholz, A., Wise, R., Wei, F., Begum, D.,

TITLE Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmens, J., Choi, D.W., Fenton, R.D., Oates, R. and Main, D., Development of a genetically and physically anchored EST resource

for barley genomics: Blumeria infected incompatible (Mla6) seedling
leaf cDNA library
Unpublished (2001)
On Jul 3, 2000 this sequence version replaced gi:8902967.

Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total bq bases = 312
Seq. primer: AATTTAACCCTCACTAAGG
High quality sequence stop: 498.

FEATURES

/organism="Hordeum vulgare"
 /cultivar="C116151 (Ma6)"
 /db_xref="taxon:4513"
 /clone="HY_CEB0006305f"
 /clone_1db="Hordeum vulgare seedling green leaf EST
 library HYCDNA0005 (Blumeria challenged)"
 /tissue_type="seedling green leaf"
 /lab_host="SOIR"
 /note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI;
 C.I. 16151 (Ma6) plants were greenhouse grown in the R
 Wise lab at Iowa State University, Ames, IA; 7 day old
 green seedlings were challenged with isolate 5874 (AvrMa6
 of Blumeria graminis f. sp. hordei), and leaves were
 harvested 20 and 24 hr post-inoculation and snap frozen;
 uninoculated leaves were harvested 20 hr post-inoculation
 (We1, Wise). In the TJ Close lab at the University of
 California, Riverside, total RNA was prepared from each
 sample pool, equal quantities of all three RNA pools were
 combined, poly(A) RNA was purified from the mixture, one
 primary unamplified cDNA library was made, and 1 million
 pfu were in vivo excised to give plus-script SK(-) cDNA
 phagemids (Choi, Close). Phagemids were plated and picked
 at the Clemson University Genomics Institute (CUGI) (Beugum
 , Palmer, Fritsch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Fritsch, Henry, Simmons, Gates

BASE COUNT
ORIGIN

130 a
199 c
396 g
156 t
2 Others

, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/Barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wang R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. *Barley Genetics Newsletter* 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgr/31/cover.html>)"

Query Match	39.0%	Score 39.4	DB 10	Length 883
Best Local Similarity	62.9%	Pred. No. 13		
Matches 61	Conservative 0	Mismatches 36	Indels 0	Gaps 0

Seq	Accession	Gene	Strain	Length
Oy	2	GCAGGTCGCGAGGGCCGAGGTGGCCGCTCCGTGGACCCACAGCGGCAGCGTGTGCATC	61	
Db	354	GCGCGCGCCGGAGCCGCGCGGGCTGGCTCCGTAGCCGCGCGCGCGCGCGGTAGCGCC	295	

Qy 62 GCCGCGCTCAACGGTCCCTACGCGCGCTGTGATGCTG 98
 ||||| ||||| ||||| ||||| |||||
 Db 294 GCCGCGCTACAGGCTGCCGTAGCCACCGCGCGCGCG 258

RESULT 10
BI778790/c

LOCUS	BT178790	528 bp	mRNA	linear	EST 23-JUL-2002
DEFINITION	EBR001_SQ001_106_R root, 3 week, hydroponic growth, no treatment, cv optic, EBR001 Hordeum vulgare cDNA clone EBR001_SQ001_106 5', mRNA sequence.				

ACCESSION	BI778790	GI:21946503
VERSION	BI778790.2	
KEYWORDS	EST.	

SOURCE ORGANISM

REFERENCE
1 (bases 1 to 528).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.

AUTHORS	Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R
TITLE	Development of Barley Transcriptome Resources

JOURNAL Unpublished (2001)
COMMENT On Sep 26, 2001 this sequence version replaced gi:15781682

Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scsi.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.

FEATURES
source

```

/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="Ebro01_SQ001_106"
/clone_1lb="root, 3 week, hydroponic grown, no treatment,
cv Optic, Ebro01"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old hydroponically grown
unsterilised barley plants. Developed as part of the barley
transcriptome resources of BBSRC/SERAD funded cereal IGF
(Investigating Gene Function) project."
151 c 215 g 79 t 2 others

```

ORIGIN

Query Match 38.8%; Score 39.2; DB 13; Length 528;
Best Local Similarity 61.4%; Pred. No. 14;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 CGGAGGCTCCGAGCGAGGTGGCCGCCGCCAGCCAGCCAGCGTGCAT 60
DB 501 CACCGCTCCCTGCGCGCGGTACACCTCCCTGCGCGCGGTGACCCCGCGNGT 442

OY 61 CGCGCGGTCAACGCTCTGACCGCCCTGTGATCGTGGCG 101
DB 441 AGCTCCGCGCACCGCGCGCTGGCGCCGATGCGCGCGCG 401

RESULT 11
BO759737/c 214 bp mRNA linear EST 26-JUL-2002
LOCUS

DEFINITION EBP105_S0004_B08_R pistil, 8 DPA, no treatment, cv Optic, EBP105
Hordeum vulgare cDNA clone EBP105_S0004_B08 5', mRNA sequence.

ACCESSION BO759737
VERSION BO759737.1 GI:21968209
KEYWORDS EST.

SOURCE Hordeum vulgare.

ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

REFERENCE 1 (bases 1 to 254)

AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

TITLE Development of Barley Transcription Resources

JOURNAL Unpublished (2001)

COMMENT Genome Dynamics/Computational Biology
Contact: Waugh R, Marshall DF

Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DU, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: estescr1.sari.ac.uk.

FEATURES
source

1..254
Location/Qualifiers

/organism="Hordeum vulgare"

/cultivar="Optic"

/db_xref="taxon:4513"

/clone="EBP105_S0004_B08"

/clone_lib="pistil", 8 DPA, no treatment, cv Optic, EBP105"

/tissue_type="pistil"

/dev_stage="8 DPA"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from pistil; dissected from developing grains (8
days post anthesis) in glasshouse grown barley plants.
developed as part of the barley transcriptome resources of
BSRC/SEERAD funded cereal IGF (Investigating gene
Function) project."

BASE COUNT 23 a 74 c 133 g 24 t

ORIGIN

Query Match 38.2%; Score 38.6; DB 14; Length 254;
Best Local Similarity 61.4%; Pred. No. 20;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 CGGAGGCTCCGAGCGAGGTGGCCGCCGCCAGCCAGCGTGCAT 60
DB 105 CACCGCTCCCTGCGCGCGGTACACCTCCCTGCGCGCGGTGACCCCGCGCGT 46

OY 61 CGCGCGGTCAACGCTCTGACCGCCCTGTGATCGTGGCG 101
DB 45 AGCTCCGCGCACCGCGCGCTGGCGCCGATGCGCGCGCG 5

RESULT 12

BO764218 349 bp mRNA linear EST 26-JUL-2002
LOCUS
DEFINITION EBP105_S0005_006_R anther, yellow stage, no treatment, cv Optic,
Eban01 Hordeum vulgare cDNA clone Eban01_S0005_006 5', mRNA
sequence.

ACCESSION BO764218
VERSION BO764218.1 GI:21972690
KEYWORDS EST.

SOURCE Hordeum vulgare.

ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

REFERENCE 1 (bases 1 to 349)

AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.

TITLE Development of Barley Transcription Resources

JOURNAL Unpublished (2001)

COMMENT Genome Dynamics/Computational Biology
Contact: Waugh R, Marshall DF

Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: estescr1.sari.ac.uk.

FEATURES
source

1..349
Location/Qualifiers

/organism="Hordeum vulgare"

/cultivar="Optic"

/db_xref="taxon:4513"

/clone="Eban01_S0005_006"

/clone_lib="anther, yellow stage, no treatment, cv Optic,
Eban01"

/tissue_type="anther"

/dev_stage="yellow stage"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from anthers dissected from developing flowers in
glasshouse grown barley plants. Developed as part of the
barley transcriptome resources of BSRC/SEERAD funded
cereal IGF (Investigating Gene Function) project."

BASE COUNT 46 a 92 c 158 g 53 t

ORIGIN

Query Match 38.2%; Score 38.6; DB 14; Length 349;
Best Local Similarity 61.4%; Pred. No. 20;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 1 CGGAGGCTCCGAGCGAGGTGGCCGCCGCCAGCCAGCGTGCAT 60
DB 325 CACCGCTCCCTGCGCGCGGTACACCTCCCTGCGCGCGGTGACCCCGCGCGT 266

OY 61 CGCGCGGTCAACGCTCTGACCGCCCTGTGATCGTGGCG 101
DB 265 AGCTCCGCGCACCGCGCGCTGGCGCCGCTGACCGCGCGCG 225

RESULT 13
BO765195/c 375 bp mRNA linear EST 26-JUL-2002
LOCUS

DEFINITION EBP103_S0006_A18_R root, 3 week, waterlogged, cv Optic, EBP103
Hordeum vulgare cDNA clone EBP103_S0006_A18 5', mRNA sequence.

ACCESSION BO765195
VERSION BO765195.1 GI:21973669
KEYWORDS EST.

SOURCE Hordeum vulgare.

ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

REFERENCE 1 (bases 1 to 375)

AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk.

FEATURES
source Location/Qualifiers
1..375

BASE COUNT 58 a 117 c 161 g 39 t
ORIGIN
Query Match 38.2%; Score 38.6; DB 14; Length 375;
Best Local Similarity 61.4%; Pred. No. 20;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 CGACGCTCCGAGCGCCGAGTGCCGCTCCGTGCGACCCACGCCGCCGAGGTGTGAT 60
Db 350 CACGCGCTCCGAGCGCCGAGTGCCGCTCCGTGCGACCCACGCCGCCGAGGTGTGAT 291
QY 61 CGCGCGGTCAACGGTCTGACGCCGCGTGCATGCGTGGC 101
Db 290 AGCCTCCGCCACCGCGCGCTGCGCGCGTGCAGCGCGCG 250

RESULT 14
LOCUS B0766713/c 402 bp mRNA linear EST 26-JUL-2002
DEFINITION EBr008_S0006_M20_R root, 3 week, drought-stressed, cv Optic,
EBr008 Hordeum vulgare cDNA clone EBr008_S0006_M20 5', mRNA
sequence.
ACCESSION B0766713
VERSION B0766713.1 GI:21975187
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 402)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk.
FEATURES
source Location/Qualifiers
1..402
/organism="Hordeum vulgare"
/cultivar="Optic"

/db_xref="taxon:4513"
/clone="EBr008_S0006_M20"
/clone_lib="root, 3 week, drought-stressed, cv Optic,
EBr008"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH108"
/note="Vector: pSPORT1; Site.1: Sal I; Site.2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old drought stressed barley
plants. Developed as part of the barley transcriptome
resources of BBSRC/SEERAD funded cereal IGF (Investigating
Gene Function) project."
BASE COUNT 60 a 127 c 169 g 46 t
ORIGIN

Query Match 38.2%; Score 38.6; DB 14; Length 402;
Best Local Similarity 61.4%; Pred. No. 20;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 CGACGCTCCGAGCGCCGAGTGCCGCTCCGTGCGACCCACGCCGCCGAGGTGTGAT 60
Db 375 CACGCGCTCCGAGCGCCGAGTGCCGCTCCGTGCGACCCACGCCGCCGAGGTGTGAT 316
QY 61 CGCGCGGTCAACGGTCTGACGCCGCGTGCATGCGTGGC 101
Db 315 AGCCTCCGCCACCGCGCGCTGCGCGCGTGCAGCGCGCG 275

RESULT 15
LOCUS BM370378/c 462 bp mRNA linear EST 23-JUL-2002
DEFINITION EBr008_S0004_A15_R root, 3 week, drought-stressed, cv Optic, EBr008
Hordeum vulgare cDNA clone EBr008_S0004_A15 5', mRNA sequence.
ACCESSION BM370378
VERSION BM370378.2 GI:21951484
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 462)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT On Jan 10, 2002 this sequence version replaced gi:18113768.
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.
FEATURES
source Location/Qualifiers
1..462
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBr008_S0004_A15"
/clone_lib="root, 3 week, drought-stressed, cv Optic,
EBr008"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH108"
/note="Vector: pSPORT1; Site.1: Sal I; Site.2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old drought stressed barley
plants. Developed as part of the barley transcriptome
resources of BBSRC/SEERAD funded cereal IGF (Investigating

BASE COUNT 53 a Gene Function) project." 88 t
ORIGIN 136 c 185 g
Query Match 38.2%; Score 38.6; DB 13; Length 462;
Best Local Similarity 61.4%; Pred. No. 19;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 CGCAGGTCGAGCGCGAGGTCGCGCTGCGGACCCACCGCGCGACGCTGCGAT 60
Db 123 CACCGCTCCCTGCGCGCGCGCTGCGGACCTGCGGCGCGCTGCGGACCGCGCGCT 64
QY 61 CGCGCGGTCAGCGGTCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 101
Db 63 AGCTCGCGCGCGCGCGCGCGCGCTGCGGCGCGCTGCGGCGCGCGCGCGCGCG 23

Search completed: November 5, 2002, 16:18:13
Job time : 1117.44 secs

APPLICANT: Sherman, D.H.

```

? APPLICANT: Liu, H.
? APPLICANT: Xue, Y.
? APPLICANT: Zhao, L.
? TITLE OF INVENTION: DNA encoding methymycin and pikromycin
? FILE REFERENCE: 600,438US1
? CURRENT APPLICATION NUMBER: US/09/0651,289
? CURRENT FILING DATE: 2001-05-18
? PRIOR APPLICATION NUMBER: 09/105,537
? PRIOR FILING DATE: 1998-06-26
? NUMBER OF SEQ ID NOS: 43
? SOFTWARE: FASTSEQ for Windows Version 3.0
? SEQ ID NO 1
? LENGTH: 15872
? TYPE: DNA
? ORGANISM: Streptomyces venezuelae
? US-09-861-289-1

```

Query Match	41.48%	Score 41.8	DB 10	Length 15872
Best Local Similarity	63.48%	Pred. No. 0.00047		
Matches 64	Conservative 0	Mismatches 37	Indels 0	Gaps 0

OY	1	CCTGGCTGGGCTGCGGGGCTTCCTGAGCCAGAGCGGCTGCATTCCTGGATGGGTTCAA	60
Db	2833	CGTCCGCCCGCTCCTCGAGTACGAGGACC	2892
OY	61	TGAGCAGGGCCCTGCACCTGTTGATGGGGGTGAGATCCGCA	101
Db	2893	GGAGCTGGGCTTCGACCTCCCTCATATTCGCTCAGGTGGCGA	2933

```

? RESULT 3
? US-09-861-289--5
? Sequence 5, Application US/09861289
? Patent No. US20020110897A1
? GENERAL INFORMATION:
? APPLICANT: Sherman, D.H.
? APPLICANT: Liu, Y.
? APPLICANT: Xue, Y.
? APPLICANT: zhaoy, L.
? TITLE OF INVENTION: DNA encoding methymycin and pikromycin
? FILE REFERENCE: 600.438US1
? CURRENT APPLICATION NUMBER: US/09/861,289
? PRIOR FILING DATE: 2001-05-18
? PRIOR APPLICATION NUMBER: 09/105,537
? PRIOR FILING DATE: 1998-06-26
? NUMBER OF SEQ ID NOS: 43
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 5
? LENGTH: 36778
? TYPE: DNA
? ORGANISM: Streptomyces venezuelae
? US-09-861-289--5

```

	Query Match	41.4%	Score 41.8	DB 10	Length 36778
	Best Local Similarity	63.4%	Pred. 0.0005		
	Matches 64	Conservative	0	Mismatches 37	Indels 0
					Gaps 0
OY	1	CCTGGCTGGGTCGTGGGGCTTCCTGCAGCCAGCCGCTCATCTGGGATGGGTTCAA	60		
Db	20049	CCTGGCTGGTGGTCTCAACACACCCCTCCCGCAGGCGCGTGCACAGGGGGGCGCTTCG	20108		
OY	61	TGAGCAGGGCCTGCACACTGTTGATGGGGGTC	3AGATCCGCA	101	
Db	20109	TGACCTGGGATTCGACTGCTGAGCGGCGGTC	3AGTCTCCGCA	20149	

RESULT 4
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.

```

1  APPLICANT: Xue, Y.
2  APPLICANT: Zhao, L.
3  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
4  FILE REFERENCE: 600,438US1
5  CURRENT APPLICATION NUMBER: US/09/661,289
6  CURRENT FILING DATE: 2001-05-18
7  PRIOR APPLICATION NUMBER: 09/109,537
8  PRIOR FILING DATE: 1998-06-26
9  NUMBER OF SEQ ID NOS: 43
10 SOFTWARE: FastSeq for Windows Version 3.0
11 SEQ ID NO 34
12 LENGTH: 4669
13 TYPE: DNA
14 ORGANISM: Streptomyces venezuelae
15 US-09-861-289-34

```

	Query Match	Score	DB	Length
Best Local Similarity	59.4%	Pred. No.	0.033	
Matches	60	Conservative	0	Mismatches
			41	Indels
			0	Gaps

	Query	DB
1	CGTGGCTGGGTCGCGGGCTTCCTGAGCCGCGCGCTCATCTCTGGATGAGGGTCAA	60
4236	CGCGCGGGCCCTACTCGGCACATCTCTCCCGACGCGGTGGCCCGCCGCGCTTCAC	4295
61	TGAGCAGGGGCTTCGACTGTTGATGAGCGGTGAGATCCGCA	101
4296	CGAGCTCGGCTTCGACTGCTGACGCGCCCTGAGCTCCGCA	4336

```

RESULT 5
US-09-925-301-474
: Sequence 474, Application us/09925301
: Patent No. US20020052308A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA106
: CURRENT APPLICATION NUMBER: US/09/925,301
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05882
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1694
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 474
: LENGTH: 3209
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (427)
: OTHER INFORMATION: n equals a,t,g, or c
: US-09-925-301-474

```

Query Match	34.1%	Score 34.4	DB 10	Length 3209
Best Local Similarity	59.0%	Pred. No. 0.064		
Matches 59	Conservative 0	Mismatches 41	Indels 0	Gaps 0

QY	1	CGTGGCTGGGTCCTGGGCTTCCTGACCCGAGCGCTCGATCTGGGATGGGCTTCA	60
DB	1225	CGTGGCACACATTCCTGGGCGATCCGGGATTCCTGGCTCTCTCAACCTGGACGCTCACTGGC	1284

QY	61	TGAGCAGGGCGCTCGACCTGTTGATGGCGGTGGAGATCCGC	100
DB	1285	GGACCTGGGGCTCGACCTGCTCATGAGCGTGGAGGTGGC	1324

```

RESULT 6
US-09-738-973-405
; Sequence 405, Application US/09738973
; Patent No. US20020110563a1
; GENERAL INFORMATION:

```

APPLICANT: Reed, Steven G.
APPLICANT: Henderson, Robert A.
APPLICANT: Lodes, Michael J.
APPLICANT: Fling, Steven P.
APPLICANT: Mohamath, Raodoh
APPLICANT: Algate, Paul A.
APPLICANT: Secrist, Heather
APPLICANT: Indrias, Carol Yoseph
APPLICANT: Benson, Darin R.
APPLICANT: Elliot, Mark
APPLICANT: Mannion, Jane
APPLICANT: Kalos, Michael D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 210121.475C9
CURRENT APPLICATION NUMBER: US/09/738,973
CURRENT FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 387
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 405
LENGTH: 435
TYPE: DNA
ORGANISM: Homo sapiens
US-09-738-973-405

Query Match 30.9%; Score 31.2; DB 10; Length 435;
Best Local Similarity 63.2%; Pred. No. 0.48; Mismatches 28; Indels 0; Gaps 0;
Matches 48; Conservative 0;

QY 1 CGTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAA 60
DB 75 CGTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAA 134

QY 61 TGAGCAGGCGCTCGAC 76
DB 135 GGAGCAGTTCTCGAC 150

RESULT 7
US-09-833-790-317
Sequence 317, Application US/09833790
Patent No. US20020068288A1
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tongtong
APPLICANT: Secrist, Heather
APPLICANT: Mohamath, Raodoh
APPLICANT: Indrias, Carol Y.
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.512
CURRENT APPLICATION NUMBER: US/09/833,790
CURRENT FILING DATE: 2001-04-11
NUMBER OF SEQ ID NOS: 440
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 317
LENGTH: 518
TYPE: DNA
ORGANISM: Homo sapiens
US-09-833-790-317

Query Match 30.9%; Score 31.2; DB 10; Length 518;
Best Local Similarity 63.2%; Pred. No. 0.49; Mismatches 28; Indels 0; Gaps 0;
Matches 48; Conservative 0;

QY 1 CGTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAA 60
DB 81 CGTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAA 140

QY 61 TGAGCAGGCGCTCGAC 76
DB 141 GGAGCAGTTCTCGAC 156

RESULT 8
US-09-861-289-30
Sequence 30, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438U51
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match 30.9%; Score 31.2; DB 10; Length 13842;
Best Local Similarity 57.0%; Pred. No. 0.65; Mismatches 43; Indels 0; Gaps 0;
Matches 57; Conservative 0;

QY 2 GTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAAT 61
DB 13249 GTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAAT 13308

QY 62 GAGCAGGCGCTCGACTCGTGTGAGGCGGTGAGATCCGCA 101
DB 13309 GACTCGGCTCGACTCCCTGACCGCGCTCGAATCCGCA 13348

RESULT 9
US-09-880-107-1542
Sequence 1542, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-MO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1542
LENGTH: 110096
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AD000092
US-09-880-107-1542

Query Match 30.9%; Score 31.2; DB 10; Length 110096;
Best Local Similarity 63.2%; Pred. No. 0.78; Mismatches 28; Indels 0; Gaps 0;
Matches 48; Conservative 0;

QY 1 CGTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAA 60
DB 85260 CGTGGCTGGGTCGTCGCTCCCTGACCGAGCGGCTCGATCCTGGAGTGGGTTCAA 85319

QY 61 TGACGAGGCGCTGCAC 76
DB 85320 GGAGCAGTTTCTGCAC 85335

RESULT 10
US-09-791-406-3
; Sequence 3, Application US/09791406
; Patent No. US20020147165A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Robert Rothlein
; APPLICANT: Takashi Kei Kishimoto
; APPLICANT: Lex M. Cowart
; TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
; FILE REFERENCE: RTS-0097
; CURRENT APPLICATION NUMBER: US/09/791,406
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; SOFTWARE: FASTSEQ for Windows Version 4.0
; TYPE: DNA
; LENGTH: 1958
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(1362)
US-09-791-406-3

Query Match 29.3%; Score 29.6; DB 10; Length 1958;
Best Local Similarity 61.8%; Pred. No. 1.6;
Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1 CGTCGTCGCGGCTGCTGCTCTCGACCGAGCGCGCTGATCTGTGGATGGGTTCAA 60
DB 120 CGTCCGTTGCTGCTGCGCTCTCGACCGTGGCCGTCGCGCGCTGCTACTTCA 179
QY 61 TGACGAGGCGCTGCAC 76
DB 180 GGAGCAGTTTCTGCAC 195

RESULT 11
US-09-927-112-3
; Sequence 3, Application US/09927112
; Patent No. US20020106774A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Silos-Santiago, Imaculada
; TITLE OF INVENTION: 32544, a novel human phospholipase C and
; FILE REFERENCE: 38155-20048.00
; CURRENT APPLICATION NUMBER: US/09/927,112
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/246,808
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; TYPE: DNA
; LENGTH: 3624
; ORGANISM: Homo sapiens
US-09-927-112-3

Query Match 28.7%; Score 29; DB 10; Length 3624;
Best Local Similarity 63.8%; Pred. No. 2.6;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 29 CCGAGCGCGCTGATCGATGGGTTCAATGAGCAGGCGCTGCATGCTGATGGCG 88
DB 2299 CCCAAGCCGCGCAGTCATGCTGGGGAGCCTTGGGGAGATCATGACCCCTTTGGGAG 2358
QY 89 GTGGAGATC 97

DB 2359 GTGGAGATC 2367

RESULT 12
US-09-927-112-1
; Sequence 1, Application US/09927112
; Patent No. US20020106774A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Silos-Santiago, Imaculada
; TITLE OF INVENTION: 32544, a novel human phospholipase C and
; FILE REFERENCE: 38155-20048.00
; CURRENT APPLICATION NUMBER: US/09/927,112
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/246,808
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; TYPE: DNA
; LENGTH: 4635
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (435)..(4058)
US-09-927-112-1

Query Match 28.7%; Score 29; DB 10; Length 4635;
Best Local Similarity 63.8%; Pred. No. 2.7;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 29 CCGAGCGCGCTGATCGATGGGTTCAATGAGCAGGCGCTGCATGCTGATGGCG 88
DB 2733 CCCAAGCCGCGCAGTCATGCTGGGGAGCCTGGGGAGATCATGACCCCTTTGGGAG 2792
QY 89 GTGGAGATC 97
DB 2793 GTGGAGATC 2801

RESULT 13
US-09-815-242-7742
; Sequence 7742, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELIPIRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 7742
;; LENGTH: 2073
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(2073)
US-09-815-242-7742

Query Match 28.5%; Score 28.8; DB 10; Length 2073;
Best Local Similarity 62.5%; Pred. No. 2.9; Mismatches 27; Indels 0; Gaps 0;
Matches 45; Conservative 0;

QY 29 CCGACGCGCTGATCTGGATGGGCTTCATGAGCAGGCGCTGATCTGGATGGCG 88
DB 1864 CAGACGGCGATCTGTCGCGCATCGTTCACAGCGCGCTGATCTGCGCTTGAATCCCA 1923
QY 89 GTGAGATCCGC 100
DB 1924 CTGCGCTTCCGC 1935

RESULT 14
US-09-954-456-1601/c
; Sequence 1601, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1601
; LENGTH: 14800
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1601

Query Match 28.3%; Score 28.6; DB 10; Length 14800;
Best Local Similarity 55.6%; Pred. No. 3.9; Mismatches 44; Indels 0; Gaps 0;
Matches 55; Conservative 0;

QY 2 GTTCCTGGGCTCTGCTCTCGACCCGAGCGGCTGATCTGGATGGGCTTCAAT 61
DB 1887 GCGGGGGAGAGCTGCGCTCGTCACTCCGCGCTGATCTGGCGCGGAATCTTC 1828

QY 62 GAGCAGGCGCTGACTGTTGATGGCGGTGAGATCCGC 100
DB 1827 GATGAGCTGTGACGCGCTCGGTGGCTGCCAGCTGCGC 1789

RESULT 15
US-09-861-289-34/c
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-34

Query Match 27.9%; Score 28.2; DB 10; Length 4689;
Best Local Similarity 59.3%; Pred. No. 4.6; Mismatches 33; Indels 0; Gaps 0;
Matches 48; Conservative 0;

QY 1 CGTCGCTGGGCTGCTGCGCTTCTCGACCCGAGCGCTGATCTGGATGGGCTTCAA 60
DB 4105 CCGCGGCGCTGCTGCGGCGCTCGGAGCGCTCGAGAGAGCTGGAGCGGACACCG 4046
QY 61 TGAGCAGGCGCTGCTGCTGCT 81
DB 4045 TGAACGCGGCGCGGACCGCT 4025

Search completed: November 5, 2002, 23:01:01
Job time : 126.254 secs

THIS PAGE BLANK (USPTO)

;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 60
DB 30562 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 30621
QY 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 101
DB 30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 30662

RESULT 3
US-09-567-969-1
;; Sequence 1, Application US/09567969
;; Patent No. 6355457
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/567,969
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 60
DB 30562 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 30621
QY 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 101
DB 30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 30662

RESULT 4
US-09-568-480-1
;; Sequence 1, Application US/09568480
;; Patent No. 6355458
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,480
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 60
DB 30562 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 30621
QY 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 101
DB 30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 30662

RESULT 5
US-09-568-486-1
;; Sequence 1, Application US/09568486
;; Patent No. 6355459
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,486
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 60
DB 30562 CGTCGCTCGGGTGGCTGCTTCTCGACCCGAGCGGCTCGATCCTGGATGGGTTCAA 30621
QY 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 101
DB 30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGAGATCCGCA 30662

RESULT 6
US-09-568-472-1
;; Sequence 1, Application US/09568472
;; Patent No. 6358719
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTGGCTGGGCTGCTGGCTTCCTCGACCGAGCGCGCTCGATCCTGGGATGGGTTCAA 60
|||
Db 30562 CGTGGCTGGGCTGCTGGCTTCCTCGACCGAGCGCGCTCGATCCTGGGATGGGTTCAA 30621

QY 61 TGACGAGGGCCTCGACTGCTGTGATGGCGGTGAGATCCGCA 101
|||

Db 30622 TGACGAGGGCCTCGACTGCTGTGATGGCGGTGAGATCCGCA 30662

RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTGGCTGGGCTGCTGGCTTCCTCGACCGAGCGCGCTCGATCCTGGGATGGGTTCAA 60
|||
Db 30562 CGTGGCTGGGCTGCTGGCTTCCTCGACCGAGCGCGCTCGATCCTGGGATGGGTTCAA 30621

QY 61 TGACGAGGGCCTCGACTGCTGTGATGGCGGTGAGATCCGCA 101
|||
Db 30622 TGACGAGGGCCTCGACTGCTGTGATGGCGGTGAGATCCGCA 30662

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:

; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.3e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGTGGCTGGGCTGCTGGCTTCCTCGACCGAGCGCGCTCGATCCTGGGATGGGTTCAA 60
|||
Db 24950 CGTGGCTGGGCTGCTGGCTTCCTCGACCGAGCGCGCTCGATCCTGGGATGGGTTCAA 25009

QY 61 TGACGAGGGCCTCGACTGCTGTGATGGCGGTGAGATCCGCA 101
|||
Db 25010 TGACGAGGGCCTCGACTGCTGTGATGGCGGTGAGATCCGCA 25050

RESULT 9
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYPEPTIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
NAME/KEY: CDS
LOCATION: 14351..19945
NAME/KEY: CDS
LOCATION: 20010..31199
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1
```

Query Match 48.3%; Score 48.8; DB 2; Length 43280;
Best Local Similarity 68.0%; Pred. No. 6.3e-06;
Matches 68; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
QY 2 GTCGCTCGGCTGCTGCTCTCTGACCCAGCGCGCTGCTGCTGGATGGGTTCAAT 61
DB 13779 GTCGCGGGCTGCTGCTGCTCTCTGAGCCCGACGCGGTGAGCCCGCGTTTCAAG 13838
QY 62 GACGAGCGCTCGACTCGTGTGATGGCGGTGAGATCCGCA 101
DB 13839 GATCTCGGCTCGACTCGTGTGATGGCGGTGAGATCCGCA 13878
```

```
RESULT 10
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafson, Claes
; APPLICANT: Ashley, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Ziemann, Brian
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144 085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3
```

Query Match 46.7%; Score 47.2; DB 4; Length 33529;
Best Local Similarity 69.6%; Pred. No. 1.8e-05;
Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
QY 5 GCTCGGCTGCTGCGCTTCTCCGACCCGAGCG3GCTGATCTGAGATGGGTTCAATGAG 64
DB 9459 GCGGCGGTGCTGAGGATCGAGGCG3GCTGAGACCGGAGCGAGGCTTTCTTGAC 9518
QY 65 CAGGCGCTCGACTGTTGATGGCGGTGAGAT 96
```

```
DB 9519 CTCGGGCTGACTGTTGATGGCGGTGAGCT 9550
```

```
RESULT 11
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438U1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-32
```

Query Match 41.4%; Score 41.8; DB 4; Length 11220;
Best Local Similarity 63.4%; Pred. No. 0.00062;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
QY 1 CGTCGCTCGGCTGCTGCTCTCTGACCCGAGCGCGCTGATCCTGGATGGGTTCAA 60
DB 4362 CCGCGCGGTGCTGCTGCTCTCTGACCCCGAGCGCGGTGAGCGGCGGCTTCG 4421
QY 61 TGAGCAGGCGCTCGACTCGTGTGATGGCGGTGAGATCCCA 101
DB 4422 TGACCTCGGATTCGACTCGTGTGAGCGGTGAGATCCCA 4462
```

```
RESULT 12
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Xue, Y.
; APPLICANT: Liu, H.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438U1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1
```

Query Match 41.4%; Score 41.8; DB 4; Length 15872;
Best Local Similarity 63.4%; Pred. No. 0.00065;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
QY 1 CGTCGCTCGGCTGCTGCTCTCTGACCCGAGCGCGCTGATCCTGGATGGGTTCAA 60
DB 2833 CGTCGCGCGCTGCTGCTCTCTGACCCGAGCGCGGTGAGCGGCGGCTTCATCA 2892
QY 61 TGAGCAGGCGCTCGACTCGTGTGATGGCGGTGAGATCCCA 101
DB 2893 GAGCTGGGCTTGCATCCCTCATGCTCGAGCTGCGGA 2933
```

RESULT 13

```

US-09-105-537-5
: Sequence 5, Application us/09105537A
: Patent No. 6265202
: GENERAL INFORMATION:
: APPLICANT: Sherman, D. H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and plikromycin
: FILE REFERENCE: 600,438051
: CURRENT APPLICATION NUMBER: US/09/105,537A
: CURRENT FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 36778
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
: US-09-105-537-5

```

Query Match	41.4%	Score 41.8	DB 4	Length 36778
Best Local Similarity	63.4%	Pred. No. 0.00073		
Matches 64	Conservative 0	Mismatches 37	Indels 0	Gaps 0
QY 1	CGTGCCTGGGGGCGCGGCTTCCTGAGCCGAGCGCGCGATCCTGGAGAGGGGTCAA	60		
Db 20049	CCTGCCCTGGTGTCTCAACACACCCTCTCCCGAGGCGCGTGCACAGCGGGGCGCTTCGG	20108		
QY 61	TGAGCAGGGCCCTGCAGCTCGTTGATGGGCGGTGGAGATTCGCA	101		
Db 20109	TGACCTGGAGTTCGACTCGTGCACGCGCGATCGAGTTCGCA	20149		

```

: RESULT 14
: US-09-320-878-19
: Sequence 19, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanje C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 30062200120
: CURRENT APPLICATION NUMBER: US/09/320, 878A
: EARLIER FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
: LENGTH: 38506
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
: IS-09-320-878-19

```

```

Query Match      41.4%; Score 41.8; DB 3; Length 38506;
Best Local Similarity 63.4%; Pred. No. 0.00073;
Matches .64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY      1 GCGCCCTGGGCGCGCGGCTTCCTGACCCGAGCGCCCTCATCTCTGGCGATGGGTTCAA 60
      |  |||||  |||||  ||  |||  ||  |||||  ||  ||||  ||  ||||

```

Db 18191 CCTCGCCGTGGTCTTCACCAACCCCTCCCCCGAGGCGCTGCACAGG66G66GCTTCGG 18250

QY 61 TGAGCAGG6GCTCGACTCGTTGATGGCGGTGAGATCCGCA 101
||| || ||||| ||| ||||| ||| |||||

Db 18251 TGACCTCGGATTGCACTCGCTACG6G6GTGACGCTCCGCA 18291

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	Similarity	41.0%	Score 41.4	DB 3	Length 80161
Best Local	Similarity	65.9%	Pred. No. 0.0011		
Matches	60	Conservative	0	Mismatches 31	Indels 0
				Gaps	0
QY	11	GTCGCGGGCTTCTCGACCCGAGCGCGTCATCTCTGGATGGGTTCAATAGCAGGCG	70		
DB	44500	GTCGCGGGACGCGAGCATGCCACGGCCCTCTGGGCCCATCGCGCGTTCAGGACTCGGA	44559		
QY	71	CTCGACTCGTTGATGGCGGTGAGATCCGA	101		
DB	44560	TTCGACTCTCTGATGACGGCTGACCTCGGCA	44590		

Search completed: November 5, 2002, 15:12:27
Job time : 189.737 secs

RESULT 1	ALIGNMENTS					
AA255887						
ID AA255887 standard; DNA: 68750 BP.						
XX AA255887;						
XX 10-APR-2000 (first entry)						
XX Sorangium cellulosum 68.75 kb contig.						
DE Sorangium cellulosum 68.75 kb contig.						
XX Epothlone biosynthesis; type I polyketide synthase; taxol substitute;						
KW anticancer; ds.						
KM Sorangium cellulosum.						
XX						
OS						
XX						
FT Key	Location/Qualifiers					
FT CDS	1..1826					
FT	/*tag- a					
FT	/partial					
FT	/product- "Partial Orf 1 protein (AAV58580)"					
FT	/note- "No initiation codon given in the specification"					
FT CDS	complement (1900..3171)					
FT	/*tag- b					
FT	/product- "Orf 2 protein (AAV58581)"					
FT CDS	3415..5556					
FT	/*tag- c					
FT	/product- "Orf 3 protein (AAV58582)"					
FT CDS	complement (5612..5992)					
FT	/*tag- d					
FT	/product- "Orf 4 protein (AAV58583)"					
FT CDS	6226..6675					
FT	/*tag- e					
FT						
10	41.8	41.4	11220	21	AA287298	S. venezuelae macr
11	41.8	41.4	15872	18	AA268715	Streptomyces venez
12	41.8	41.4	15872	21	AA287283	S. venezuelae vep
13	41.8	41.4	36778	21	AA287318	S. venezuelae pik
14	41.8	41.4	37948	21	AA287285	Nucleotide sequenc
15	41.8	41.4	38506	21	AA275633	Nucleotide sequenc
16	41.8	41.4	38506	21	AA256001	Recombinant cosmid
17	41.4	41.0	990	21	AA255784	Type I polyketide
18	41.4	41.0	9513	22	AA288337	S. spinosa DNA fra
19	41.4	41.0	12249	21	AA255840	Complete Mitomycin
20	41.4	41.0	18331	21	AA255857	Complete nucleotid
21	41.4	41.0	50000	22	AA288313	S. spinosa DNA fra
22	41.4	41.0	50000	22	AA288316	S. spinosa DNA fra
23	41.4	41.0	80161	20	AA221501	DNA fragment of Sa
24	40.2	39.8	77536	21	AA214651	Nucleotide sequenc
25	40.2	39.8	109519	22	AA208693	Micromonospora t
26	40.2	39.8	4403765	22	AA208693	Mycobacterium tub
27	40	39.6	7788	22	AA288335	S. spinosa DNA fra
28	40	39.6	50000	22	AA288312	S. spinosa DNA fra
29	39.2	38.8	30690	21	AAA92301	S. avermitilis ave
30	39.2	38.8	30690	22	AA279277	Streptomyces averm
31	38.8	38.4	29879	14	AA246806	erya region of S.
32	38.2	37.8	77536	21	AA214651	Nucleotide sequenc
33	37.6	37.2	47981	22	AA230757	Micromonospora me
34	37.4	37.0	28958	17	AA206769	Sorangium cellulos
35	37.4	37.0	28958	18	AA289956	Sorangium cellulos
36	37.4	37.0	28958	21	AA275299	DNA sequence of So
37	37.4	37.0	49377	19	AA205287	The soraphen biosy
38	36.6	36.2	3978	21	AA255785	Type I polyketide
39	36	35.6	5088	22	AA290038	Nucleotide sequenc
40	36	35.6	6459	22	AA288336	S. spinosa DNA fra
41	36	35.6	34071	22	AA290033	Nucleotide sequenc
42	36	35.6	42717	22	AA290032	Nucleotide sequenc
43	35.4	35.0	1681	24	AA218438	Contig 115 DNA enc
44	35.4	35.0	4689	21	AA287299	S. venezuelae macr
45	35	34.7	16767	22	AA288339	S. spinosa DNA fra

FT	/product= "Orf 5 protein (AAV58584)"
FT	7610..11875
FT	//tag- f
FT	/product= "Type I polyketide synthase, EPOS A
FT	(AAV58573)"
FT	11872..116104
FT	//tag- g
FT	/product= "Non-ribosomal peptide synthetase, EPOS P
FT	(AAV58574)"
FT	16251..21749
FT	//tag- h
FT	/product= "Type I polyketide synthase, EPOS B
FT	(AAV58575)"
FT	21746..43519
FT	//tag- i
FT	/product= "Type I polyketide synthase, EPOS C
FT	(AAV58576)"
FT	43524..54920
FT	//tag- j
FT	/product= "Type I polyketide synthase, EPOS D
FT	(AAV58577)"
FT	54935..62254
FT	//tag- k
FT	/product= "Type I polyketide synthase, EPOS E
FT	(AAV58578)"
FT	62369..63628
FT	//tag- l
FT	/product= "Cytochrome P450 oxygenase homologue, EPOS F
FT	(AAV38579)"
FT	63779..64333
FT	//tag- m
FT	/product= "Orf 6 p:protein (AAV58585)"
FT	complement (63853..64290)
FT	//tag- n
FT	/product= "Orf 7 p:protein (AAV58586)"
FT	64363..64920
FT	//tag- o
FT	/product= "Orf 8 protein (AAV58587)"
FT	complement (64287..64727)
FT	//tag- p
FT	/product= "Orf 9 protein (AAV58588)"
FT	65063..65767
FT	//tag- q
FT	/product= "Orf 10 protein"
FT	complement (65008..65874)
FT	//tag- r
FT	/product= "Orf 11 Protein (AAV58590)"
FT	complement (65871..66338)
FT	//tag- s
FT	/product= "Orf 12 protein (AAV58591)"
FT	66667..67137
FT	//tag- t
FT	/product= "Orf 13 p:protein (AAV58592)"
FT	67334..68251
FT	//tag- u
FT	/product= "Orf 14 p:protein (AAV58593)"
FT	68346..68750
FT	/partial v
FT	//tag- v
FT	/product= "Partial Orf 15 protein (AAV58594)"
FT	/note= "No termination codon given in the specification
XX	
PN	W09966028-AZ.
XX	
PD	23-DEC-1999.
XX	
PE	16-JUN-1999;
XX	
PR	18-JUN-1998; 98US-0099504.
XX	
PR	24-SEP-1998; 98US-0101631.
XX	
PR	05-FEB-1999; 99US-0118906.
XX	
PA	(NOVS) NOVARTIS AG.

PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PI	Schupp T, Ligon JM, Molnar I, Zittle R, Goerlach J, Cyr D;
XX	
DR	WPI: 2000-097741/08
DR	P-PSDB: AAY585773, AAY58674, AAY58675, AAY58676, AAY58677, AAY58678,
DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
XX	AAY58592, AAY58593, AAY58594.
PT	New isolated epoethione synthase genes, used for the recombinant
PT	production of epoethione for use in cancer therapy
XX	
PS	Claim 14; Page 87-104; 174pp; English.
XX	
CC	This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC	comprising 22 open reading frames (ORFs) and includes genes encoding
CC	proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC	starter unit; polyketides being synthesised from two-carbon building
CC	blocks, the beta-carbon of which always carries a keto group. Each round
CC	of two-carbon addition is carried out by a complex of enzymes known as
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC	EPoS A (AAY58573) and EPoS P (AAY58574) are involved in formation of
CC	the thiazole ring formation of epoethiones, and EPoS B, EPoS C, EPoS D
CC	and EPoS E (AAY58575-58578) are involved in polyketide backbone
CC	formation. EPo F (AAY58579) is an epoethione macrolactone oxidase, and
CC	the proteins Ori 3 (AAY58582) and Ori14 (AAY58593) are thought to be
CC	involved in transport. Epoethiones mimic the biological activity of
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic
CC	compositions. Epoethiones exhibit a much lower drop in potency against a
CC	multiply drug-resistant cell line compared with taxol, and are
CC	considerably less efficient exported from such cells by the multidrug
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC	epoethiones as anticancer agents, P-glycoprotein) are problematical to produce on a
CC	large scale. Epoethiones are too complex for industrial scale chemical
CC	synthesis, and Sorangium cellulosum is difficult to ferment producing
CC	poor yields of epoethiones. The nucleic acids of the invention may be
CC	used for the recombinant production of epoethiones in a heterologous host
CC	that is more amenable to fermentation.
XX	
SQ	Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
	Query Match 100.0%; Score 101; DB 21; Length 68750;
	Best Local Similarity 100.0%; Pred. No. 1.5e-20;
	Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 CGTGCCTGGGCGTCTGACCTCTCGACCGCGCCGCTGATTCCTGGATGGCGTTCAA 60
DB	30562 CGTGCCTGGGCGTCTGACCTCTCGACCGCGCGGCGATTCCTGGATGGCGTTCAA 30621
OY	61 TGAGCAGGCGCTCGACTCGTTGATGGCGGTGAGATCCGCA 101
DB	30622 TGAGCAGGCGCTCGACTCGTTGATGGCGGTGAGATCCGCA 30662
RESULT 2	
ID	AAA29349
AC	AAA29349 standard; DNA; 71989 BP.
XX	
AC	AAA29349;
DT	12-SEP-2000 (first entry)
XX	
DE	Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
KM	Epoethione polyketide synthase; epok; epob; epoc; epod; epof; epog;
KM	epok; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
XX	tubulin polymerization assay; anti-tumour; cytostatic; ds.
OS	Sorangium cellulosum.
XX	
Key	Location/Qualifiers
XX	

FT	CDS	3..992	/*tag= t
FT		/label= a	/note= "encodes adenylation domain A8 of the NRPS module"
FT		/product= transposase	9183..9992
FT		/note= "not part of the PKS"	/*tag= u
FT	CDS	989..1501	/note= "encodes oxidation region for forming thiazole"
FT		/*tag= b	10121..10138
FT		/label= ORF_B	/*tag= v
FT		/product= transposase	/note= "encodes adenylation domain A10 of the NRPS module"
FT		/note= "not part of the PKS"	10261..10306
FT	CDS	1998..6263	/*tag= w
FT		/*tag= c	/note= "encodes thiolation domain (PCP) of the NRPS module"
FT		/label= epoc_gene	10639..16137
FT		/note= "encodes the loading domain"	/*tag= x
FT	misc_RNA	2031..3548	/label= epoc_gene
FT		/*tag= d	/note= "encodes module 2"
FT		/note= "encodes ketide synthase (KS-Y) of the loading domain"	10654..12033
FT		3621..4661	/*tag= y
FT		/*tag= e	/note= "encodes KS2, the KS domain of module 2"
FT		/note= "encodes acyl transferase (AT) of the loading domain"	12250..13287
FT		4917..5810	/*tag= z
FT	misc_RNA	/*tag= f	/note= "encodes AT2, the AT domain of module 2"
FT		/note= "encodes enoyl reductase (ER) of the loading domain, potentially involved in formation of the thiazole moiety"	13327..13899
FT		5856..6155	/*tag= aa
FT		/*tag= g	/note= "encodes dehydratase (DH) 2, the DH domain of module 2"
FT		/note= "encodes acyl carrier protein (ACP) of the loading domain"	14962..15756
FT	CDS	6260..10493	/*tag= ab
FT		/*tag= h	/note= "encodes ketoreductase (KR) 2, the KR domain of module 2"
FT		/label= epob_gene	15763..16008
FT		/note= "encodes module 1, the NRPS module"	/*tag= ac
FT	misc_RNA	2031..3548	/note= "encodes ACP2, the ACP domain of module 2"
FT		/*tag= i	16134..37907
FT		/note= "encodes condensation domain C2 of the NRPS module"	/*tag= ad
FT		2031..3548	/label= epod_gene
FT		/*tag= j	/note= "encodes modules 3-6"
FT		2031..3548	/*tag= ae
FT		/*tag= k	/note= "encodes KS3"
FT	misc_RNA	6861..6887	17817..18857
FT		/note= "encodes condensation domain C2 of the NRPS module"	/*tag= af
FT		6861..6887	/note= "encodes AT3"
FT		/*tag= l	19581..20396
FT		/note= "encodes condensation domain C4 of the NRPS module"	/*tag= ag
FT		7358..7366	/note= "encodes KR3"
FT		/*tag= m	20424..20642
FT		/note= "encodes condensation domain C7 (partial) of the NRPS module"	/*tag= ah
FT		7898..7921	/note= "encodes ACP3"
FT		/*tag= n	20706..22082
FT		/note= "encodes adenylation domain A1 of the NRPS module"	/*tag= ai
FT		7898..7921	/note= "encodes KS4"
FT		/*tag= o	22296..23336
FT		/note= "encodes adenylation domain A1 of the NRPS module"	/*tag= aj
FT		8261..8308	/note= "encodes AT4"
FT		/*tag= p	24059..24647
FT		/note= "encodes adenylation domain A3 of the NRPS module"	/*tag= ak
FT		8411..8422	/note= "encodes KR4"
FT		/*tag= q	24867..25151
FT		/note= "encodes adenylation domain A4 of the NRPS module"	/*tag= al
FT		8861..8905	/note= "encodes ACP4"
FT		/*tag= r	25203..26576
FT		/note= "encodes adenylation domain A6 of the NRPS module"	/*tag= am
FT		8966..8983	/note= "encodes KS5"
FT		/*tag= s	26793..27883
FT		/note= "encodes adenylation domain A7 of the NRPS module"	/*tag= an
FT		9090..9179	/note= "encodes AT5"
FT		/*tag= t	27966..28574
FT		/*tag= u	/note= "encodes DH5"
FT		/*tag= v	29433..30287
FT		/*tag= w	/*tag= ap

```
FT      misc_RNA      /note- "encodes E35"
FT      30321..30869   /tag- aq
FT      /note- "encodes K35"
FT      31077..31373   /tag- ar
FT      /note- "encodes AK:P5"
FT      31440..32807   /tag- as
FT      /note- "encodes K3:6"
FT      33018..34067   /tag- at
FT      /note- "encodes A3:6"
FT      34107..34676   /tag- au
FT      /note- "encodes D3:6"
FT      35760..36641   /tag- av
FT      /note- "encodes E3:6"
FT      36705..37256   /tag- aw
FT      /note- "encodes K3:6"
FT      37470..37769   /tag- ax
FT      /note- "encodes AC3:6"
FT      37912..49308   /tag- ay
FT      /label- epof-gene
FT      /note- "encodes modules 7 and 8"
FT      38014..39375   /tag- az
FT      /note- "encodes KS1"
FT      39589..40626   /tag- ba
FT      /note- "encodes AT1"
FT      41341..41922   /tag- db
FT      /note- "encodes KR1"
FT      42181..42423
FT      misc_RNA      /note- "encodes KR1"
FT      42181..42423

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.5e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CGTCGCTGGGCTGCTGGGCTTCTCGACCCGAGCGCGCTCGATCTGGGATGGGCTTCAA 60
DB      24950 CGTCGCTGGGCTGCTGGGCTTCTCGACCCGAGCGCGCTCGATCTGGGATGGGCTTCAA 25009
QY      61 TGAGCAGGCGCTCGACTCGTTGATGGCGGT3GAGATCCGCA 101
DB      25010 TGAGCAGGCGCTCGACTCGTTGATGGCGGT 3GAGATCCGCA 25050

RESULT 3
AAT80413
ID      AAT80413 standard; DNA: 43280 BP.
AC      AAT80413;
XX      27-FEB-1998 (first entry)
XX      Ty lactone synthase gene cluster.
DE      Ty lactone synthase gene cluster.
KW      Ty lactone synthase gene cluster; ty1G gene; multifunctional protein;
XX      polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
OS      Streptomyces fradiae.
XX      key      Location/Qualifiers
FH      CDS      816..14243
FT      /tag- a
FT      /transl_except- (pos: 816..818, aa: Met)
FT      /note- "ORF1 encodes protein shown in AAW22601"
```

```
FT      CDS      14351..19945
FT      /tag- b
FT      /transl_except- (pos: 14351..14353, aa: Met)
FT      /note- "ORF2 encodes protein shown in AAW22602"
FT      20010..31199
FT      CDS      /tag- c
FT      /transl_except- (pos: 20010..20012, aa: Met)
FT      /note- "ORF3 encodes protein shown in AAW22603"
FT      31232..36067
FT      CDS      /tag- d
FT      /note- "ORF4 encodes protein shown in AAW22604"
FT      36249..41774
FT      CDS      /tag- e
FT      /note- "ORF5 encodes protein shown in AAW22605"

FT      PN      EF791655-A2.
XX      27-AUG-1997.
PD      19-FEB-1997; 97EP-0301056.
XX      22-FEB-1996; 96US-0012078.
PR      (ELI ) LILIX & CO ELI.
PA      Dehoff BS, Kuhstoss SA, Rostack PR, Sutton KL;
XX      WPI: 1997-418046/39.
XX      P-PADB: AAW22601-W22605.
DR      DNA encoding Streptomyces fradiae ty lactone synthase domain - for
XX      production of tylosin-related polyketide compounds
XX      Claim 2: Pages 8-66; 220pp; English.
PS      This sequence represents the ty lactone synthase gene cluster of the
XX      invention. This sequence is also referred to as the ty1G gene, and was
XX      isolated from Streptomyces fradiae. This sequence encodes multifunctional
XX      proteins which direct the synthesis of the polyketide ty lactone, isolated
XX      from Streptomyces fradiae. Ty lactone is the basic building block of the
XX      antibiotic tylosin. The DNA sequence can be modified so as to alter the
XX      type of carboxylic acids incorporated, the number of carboxylic acids
XX      incorporated and/or the post-condensation reactions performed, thereby
XX      resulting in novel tylosin-related polyketides.
XX      Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
SQ

Query Match      48.3%; Score 48.8; DB 18; Length 43280;
Best Local Similarity 68.0%; Pred. No. 4.1e-05;
Matches 68; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      2 GTCGCTGGGCTGCTGGGCTTCTCGACCCGAGCGCGCTCGATCTGGGATGGGCTTCAAT 61
DB      13779 GTCGCTGGGCTGCTGGGCTTCTCGACCCGAGCGCGCTCGATCTGGGATGGGCTTCAAG 13838
QY      62 GAGCAGGCGCTCGACTCGTTGATGGCGGTGAGATCCGCA 101
DB      13839 GATCTCGGCTTCGACTCGTGCACCGCGCTGAGACTCCGCA 13878

RESULT 4
AAS17367
ID      AAS17367 standard; DNA: 33529 BP.
AC      AAS17367;
XX      12-MAR-2002 (first entry)
XX      DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
DE      Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
XX      acyl transferase domain; dehydratase domain; ketoreductase domain;
XX      acyl carrier protein domain; pKOS28-26; ds.
KW
```


XX OS Sorangium cellulosum.
XX PN US6280999-B1.
XX PD 28-AUG-2001.
XX PF 31-AUG-1998; 98US-0144085.
XX PR 22-JAN-1998; 98US-0010809.
XX PA (KOSA-) KOSAN BIOSCIENCE.
XX PI Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX DR WPL; 2001-606536/69.
XX PT Novel purified, isolated DNA molecule from Sorangium cellulosum having
XX PT polyketide open reading frame encoding modules with one or more domains
XX PT such as ketosynthase, acyl transferase and acyl carrier protein domains
XX PT
XX PS Claim 4; Fig 1; 72pp; English.
XX CC The present invention relates to the isolation of novel Sorangium
XX CC cellulosum polyketide synthases (PKS), and the polynucleotide sequences
XX CC encoding them. The polyketide synthases include catalytic domains such
XX CC as ketosynthase domain, acyl transferase domain, dehydratase domain,
XX CC ketoreductase domain and acyl carrier protein domain. A host cell
XX CC comprising a PKS ORF (open reading frame) which encodes one or more
XX CC more PKS domains is useful for producing polyketide synthases from which
XX CC polyketides can be produced. The host cells are useful for constructing
XX CC a library, where each individual colony of the library represents a
XX CC colony with the ability to produce a particular PKS synthase and
XX CC ultimately a particular polyketide. The polyketides produced by these
XX CC colonies can be used collectively in a panel to represent a library or
XX CC may be assessed individually for activity. Colonies in the library are
XX CC also induced to produce the relevant synthases and thus to produce the
XX CC relevant polyketides to obtain a library of candidate polyketides which
XX CC can be screened for binding to desired targets such as receptors,
XX CC signalling proteins, etc. The present sequence represents the DNA
XX CC sequence of cosmid pK0528-26 which encodes one or more domains of
XX CC S. cellulosum PKS.
XX CC Note: The present sequence is said to encode the functional domains
XX CC of S. cellulosum PKS which correspond to domains or domain subsets of
XX CC the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX CC
XX SQ Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

Query Match 46.7%; Score 47.2; DB 23; Length 33529;
Best Local Similarity 69.6%; Pred. No. 0.00012;
Matches 64; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 5 GCTCGGTCGTCGTCGACCGCGCTCGATCCGCGGATGGGTTCAATGAG 64
DB 9459 GCGGGGTCGTCGTCGACCGCGCTCGATCCGCGGATGGGTTCAATGAG 64
QY 65 CAGGGCTCGATCGTTGATGCGCGGTGAGAT 96
DB 9519 CTCGGGTCGACTCGTTGATGCGGTGAGCT 9550

RESULT 5
AADI184
ID AADI184 standard; DNA; 65140 BP.
AC AADI184;
XX
XX 29-NOV-2001 (first entry)
DE Streptomycetes noursei nys1 DNA of nystatin PKS gene cluster.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

KW antifungal; antibiotic; nys1; ds.
XX OS Streptomycetes noursei.
XX Key Location/Qualifiers
XX CDS complement (1..1035)
XX FT /tag= a
XX FT /product= "NysD2 partial protein"
XX FT /note= "CDS does not include stop codon"
XX FT complement (1056..2576)
XX FT /tag= b
XX FT /product= "NysD1 protein"
XX FT 2806..6906
XX FT /tag= c
XX FT /product= "NysA protein"
XX FT 6952..16530
XX FT /tag= d
XX FT /product= "NysB protein"
XX FT 16550..49840
XX FT /tag= e
XX FT /product= "NysC protein"
XX FT 50260..51015
XX FT /tag= f
XX FT /product= "NysE protein"
XX FT 51405..54305
XX FT /tag= g
XX FT /product= "NysR1 protein"
XX FT 54329..57190
XX FT /tag= h
XX FT /product= "NysR2 protein"
XX FT /note= "CDS does not include start codon"
XX FT 57180..59963
XX FT /tag= i
XX FT /product= "NysR3 protein"
XX FT 60415..61047
XX FT /tag= j
XX FT /product= "NysR4 (short) protein"
XX FT /note= "CDS does not include start codon"
XX FT 61736..62497
XX FT /tag= k
XX FT /product= "NysR5 protein"
XX FT /note= "CDS does not include start codon"
XX FT complement (62551..63615)
XX FT /tag= l
XX FT /product= "ORF2 protein"
XX FT /note= "CDS does not include start codon"
XX FT 63765..64961
XX FT /tag= m
XX FT /product= "ORF1 protein"

W0200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SMTF-) SINTER STIFTTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (FJAE/) FJAEVYIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Walla S, Ellingsen TE, Sletta H, Gulliksen O;

```

XX MO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX
XX 10-APR-2000; 2000GB-0008786.
XX
XX 14-APR-2000; 2000GB-0009387.
XX
PA (UNIV-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SMTF) SINTEF STEIPELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINEVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (ETAE/) FJÆRVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
PA
PI Zotchev SB, Sekurova ON, Fjærvik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI: 2001-557614/62.
XX
XX DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX
XX DR AAE10149, AAE10150.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX PT useful as antibiotics and antifungals -
XX
XX PS Claim 1; Page 188-254; 266pp; English.
XX
XX CC The present invention relates to the cloning and sequencing of the gene
XX CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
XX CC The nystatin PKS is useful as antifungal antibiotics. The present
XX CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX SO Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other:
XX
XX Query Match 45.1%; Score 45.6; DB 22; Length 125401;
XX Best Local Similarity 66.0%; Pred. No. 0.00041;
XX Matches 66; Conservative 0; Mismatches 34; Indels 0; Caps 0;
XX
XX QY 2 GTCCGTCGGGCTCGTGCTTCCTGCACGCCGAGCGCGCTCGATCTGTGGATGGGTTCAAT 61
XX | ||| ||||| ||| | |||| | ||| ||| ||| ||| ||| |
Db 109649 GCCCGCGGTGCTGCTGCGGACGCCGACGCGCTGCGGCGATGCGGGGCGACAGTTCAG 109708
XX ||| ||||| ||||| ||| ||||| ||||| ||| ||||| |||||
XX Db 109709 GAGTTGGGCTTCGACTCGCTGACCGCGGTGGAGTTCCGCA 109748
XX
XX RESULT 7
XX AAF24892
XX ID AAF24892 standard; DNA; 20394 BP.
XX
XX AAF24892:
XX AC
XX DT 20-APR-2001 (first entry)
XX DE Pimaricin biosynthesis associated polyketide synthase gene.
XX KW Polyketide synthase; oxidative modification; metabolite; antibiotic;
XX KW anticancer; pimaricin; ss.
XX OS Streptomyces natalensis.
XX
XX Key Location/Qualifiers
XX CDS 1..20394
XX /tag= a
XX FT

```

FT /product= "polyketide synthase"
XX
XX WO20007222-A1.
XX
XX 21-DEC-2000.
XX
XX 14-JUN-2000; 2000WO-EP06227.
XX
XX 14-JUN-1999; 99EP-0201893.
XX
XX (STAM) DSM NV.
XX
XX Martin JF, Aparicio JF, Colina AJ;
XX
XX WPI: 2001-080693/09.
XX
XX P-PSDB: AAB31558.
XX
XX New polynucleotides encoding enzymes involved in the biosynthesis of
XX
XX pimaricin, useful for modifying the biosynthesis of pimaricin and in
XX
XX the synthesis of new compounds -
XX
XX
XX Disclosure; Page 53-80; 116pp; English.
XX
XX The present sequence encodes a polyketide synthase which is associated
XX
XX with the biosynthesis of pimaricin. The polyketide synthase polypeptide
XX
XX is useful for the oxidative modification of a methyl group of a suitable
XX
XX compound, e.g. a bioactive compound including a secondary metabolite,
XX
XX antibiotics and anticancer agents. Recombinant cells comprising the
XX
XX gene are useful for the production of pimaricin. The polyketide synthase
XX
XX polynucleotide may be over expressed in Streptomyces, leading to an
XX
XX increase in the biosynthesis of pimaricin, as a source of primers for
XX
XX amplification reaction and as probes.
XX
XX Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;
SQ
Query Match 42.6%; Score 43; DB 22; Length 20394;
Best Local Similarity 64.6%; Pred. No. 0.0019;
Matches 64; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
OY 3 TCGCTCGGGTCTGGGCTTCCTCGACCCGAGCGCGCTGCATCCTGGATGGGCTTCAGT 62
DB 14738 TCGCAGACGTCCTCGGCGACCGCGCGCGGAAACGATGCCCTGACCGGGCTTCAGG 14797
OY 63 AGCAGGCGCTGCATCGTTGATGGCGGTGAGATCCGCA 101
DB 14798 ACCTGGGCTGCACTCCTGAGCGGATCGAACTCCGTA 14836
RESULT 8
AAV21186
ID AAV21186 standard; DNA; 5676 BP.
XX
XX AAV21186;
XX
XX 24-JUL-1998 (first entry)
XX
XX Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.
XX
XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX
XX polyketide synthase; actinomycete; ansamycin; ds.
XX
XX Amycolatopsis mediterranei.
XX
XX
XX Key Location/Qualifiers
XX
XX CDS 3..5676
XX
XX /*tag= a
XX
XX /product= "polyketide synthase"
XX
XX /note= "no stop codon given"
XX
XX
XX WO9807868-A1.
XX
XX 26-FEB-1998.
XX
XX

PF 18-AUG-1997; 97WO-EP04495.
XX
XX 20-AUG-1996; 96EP-0810551.
XX
XX (NOVS) NOVARTIS AG.
XX
XX Engel N, Schupp T, Toupet C;
XX
XX WPI: 1998-169172/15.
XX
XX P-PSDB: AAM52844.
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX
XX to produce rifamycin and rifamycin analogues
XX
XX
XX Claim 4; Page 38-43; 205pp; English.
XX
XX The present sequence represents Amycolatopsis mediterranei strain wt3136
XX
XX 5.7 kb KpnI fragment DNA, from the present invention. The present
XX
XX invention describes a Amycolatopsis mediterranei rifamycin synthesis
XX
XX gene cluster DNA fragment comprising a DNA region involved directly or
XX
XX indirectly in the gene cluster responsible for rifamycin synthesis,
XX
XX including the adjacent DNA regions to the right and left which, by
XX
XX reason of their function in connection with rifamycin biosynthesis,
XX
XX qualify as constituents of this rifamycin gene cluster, and functional
XX
XX fragments, derivatives or constituents of these. The Amycolatopsis
XX
XX mediterranei rifamycin synthesis gene cluster DNA fragment can be used
XX
XX for producing rifamycin, rifamycin analogues or precursors. It can also
XX
XX be used for inactivating or modifying genes involved in ansamycin or
XX
XX rifamycin biosynthesis. The DNA can be used for constructing mutant
XX
XX actinomycetes strains from which the natural rifamycin or ansamycin
XX
XX biosynthesis gene cluster has been partly or completely deleted. The
XX
XX DNA fragment can be used for assembling a library of polyketide
XX
XX synthases, which can be used for assembling a library of polyketides.
XX
XX A hybridisation probe of the invention can be used for identifying DNA
XX
XX fragments involved in the biosynthesis of ansamycins.
XX
XX Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;
SQ
Query Match 42.0%; Score 42.4; DB 19; Length 5676;
Best Local Similarity 64.0%; Pred. No. 0.0025;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY 2 GTGCTCGGGTCTGGGCTTCCTCGACCCGAGCGCTGCATCCTGGATGGGCTTCAGT 61
DB 3132 GTGCGGCTGTGTCCTGGGACCGCGCGCGCGCGCTGCCGCGGACAGCGGCTTCAG 3191
OY 62 GAGCAGGCGCTGCATCGTTGATGGCGGTGAGATCCGCA 101
DB 3192 GACACCGGCTTGCATCGCTGACGTGGTGGAGATCGCA 3231
RESULT 9
AAV21187
ID AAV21187 standard; DNA; 53789 BP.
XX
XX AAV21187;
XX
XX 24-JUL-1998 (first entry)
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX
XX polyketide synthase; actinomycete; ansamycin; ds.
XX
XX Amycolatopsis mediterranei.
XX
XX
XX Key Location/Qualifiers
XX
XX CDS 1825..15543
XX
XX /*tag= a
XX
XX /label= ORF_A
XX
XX /product= "polyketide synthase"
XX
XX 15550..30759
XX
XX /*tag= b
XX
XX
XX CDS
XX
XX
XX

FT	/label- ORF_B	/product= "polyketide synthase"
FT	30895..36060	
FT	/tag- C	
FT	/label- ORF_C	
FT	/product= "polyketide synthase"	
FT	36259..41325	
FT	/tag- d	
FT	/label- ORF_D	
FT	/product= "polyketide synthase"	
FT	41373..51614	
FT	/tag- e	
FT	/label- ORF_E	
FT	/product= "polyketide synthase"	
FT	51713..5293	
FT	/tag- f	
FT	/label- ORF_F	
FT	/product= "polyketide synthase"	
FT		
XX	'W09807868-A1.	
PN		
PD	26-FEB-1998.	
XX		
XX	18-AUG-1997:	97WO-EP04495.
XX		
XX	20-AUG-1996:	96EP-0810531.
XX		
XX	(NOVS) NOVARTIS AG.	
XX		
PI	Engel N., Schupp T., Toupet C;	
XX		
XX	WPI: 1998-169172/15.	
DR	P-PSDB: AAM52845-W52850.	
XX		
PT	Amycolatopsis mediterranei rifamycin synthesis gene cluster - used	
PT	to produce rifamycin and rifamycin analogues	
XX		
PS	Claim 4; Page 53-102; 205pp; English.	
XX		
CC	The present sequence represents a Amycolatopsis mediterranei rifamycin	
CC	synthesis gene cluster DNA fragment from the present invention. The	
CC	DNA fragment comprises a DNA region involved directly or indirectly	
CC	in the gene cluster responsible for rifamycin synthesis, including	
CC	the adjacent DNA regions to the right and left which, by reason of	
CC	their function in connection with rifamycin biosynthesis, qualify	
CC	as constituents of this rifamycin gene cluster, and functional	
CC	fragments, derivatives or constituents of these. The Amycolatopsis	
CC	mediterranei rifamycin synthesis gene cluster DNA fragment can be used	
CC	for producing rifamycin, rifamycin analogues or precursors. It can also	
CC	be used for inactivating or modifying genes involved in ansamycin or	
CC	rifamycin biosynthesis. The DNA can be used for constructing mutant	
CC	actinomycetes strains from which the natural rifamycin or ansamycin	
CC	biosynthesis gene cluster has been partly or completely deleted. The	
CC	DNA fragment can be used for assembling a library of polyketide	
CC	synthases, which can be used for assembling a library of polyketides.	
CC	A hybridisation probe of the invention can be used for identifying DNA	
CC	fragments involved in the biosynthesis of ansamycins.	
XX		
XX		
SO	Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;	
	Query Match	42.0% Score 42.4; DB 19; Length 53789;
	Best Local Similarity	64.0% Pred. IC: 0.0033;
Matches	64; Conservative	0; Mismatches 36; Indels 0; Gaps 0
OY	2	GTTCGCTGGGCTCTGGGCTTCTCGACCCGAGCGCGCTGATCTGGGATGGGGTTCAAT 61
DB	25291	GTTCGCGCCGCTCTCGGGGACCCGAGCGGAGAGCGGCTCGCGCTGACACGGGCTTCAAG 25350
OY	62	GAGCAGGGGCTTCGACTGTTGATGGCGGTGAGATCCGA 101
DB	25351	GACGCGGCTTCGACTGCTGACCGCGGCTGAGCTCGGCGCA 25390

RESULT 10	
ID	AAZ87298 standard; DNA: 11220 BP.
AC	AAZ87298
XX	
AC	AAZ87298;
XX	
DT	05-JUN-2000 (first entry)
XX	
DE	S. venezuelae macrolide biosynthetic gene plkAII, SEQ ID NO:32.
XX	
KM	Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KM	neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM	chronic obstructive pulmonary disease; respiratory inflammation;
KM	hypercholesterolaemia; crop protection agent; ds.
XX	
OS	Streptomyces venezuelae ATCC15439.
XX	
FT	Key Location/Qualifiers
FT	CDS 1..11220
FT	/*tag= "a
FT	/product= "PlkAII"
XX	
PN	WO200000620-A2.
XX	
PD	06-JAN-2000.
XX	
PF	25-JUN-1999; 99WO-US14398.
XX	
PR	26-JUN-1998; 98US-0105537.
XX	
PA	(MINU) UNIV MINNESOTA.
XX	
PI	Sherman DH, Liu H, Xue Y, Zhao L;
XX	
DR	WPI: 2000-160679/14.
XX	
P-PSDB:	AAV771193.
XX	
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX	
PT	synthesis of methymycin and pikromycin -
XX	
PS	Claim 15; Page 398-403; 438pp; English.
XX	
CC	The invention relates to an isolated and purified nucleic acid segment
CC	comprising a desosamine biosynthetic gene cluster, a fragment or its
CC	biologically active variant, where the nucleic acid sequence is not
CC	derived from the eryc gene cluster of <i>Saccharopolyspora erythraea</i> or
CC	<i>Streptomyces antibioticus</i> .
CC	biosynthetic gene cluster, or fragments thereof. The macrolide
CC	biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC	pikromycin, neomethymycin, narbomycin or a combination of these
CC	compounds. Recombinant or augmented cells comprising the desosamine
CC	and/or macrolide biosynthetic gene clusters are useful for the
CC	production of biologically active macrolides. The macrolide biosynthetic
CC	proteins are useful for synthesis of methymycin, pikromycin,
CC	neomethymycin and narbomycin. The alternative termination of polyketide
CC	synthesis may be useful to prepare novel antibiotics and
CC	polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC	recombinant host cells are useful as biopolymers, e.g., in packaging or
CC	biomedical applications, to engineer PHA monomer synthases or to prepare
CC	immunologically active agents, such as chemotherapeutics,
CC	immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC	disease as well as other diseases involving respiratory inflammation,
CC	cholesterol-lowering agents or macrolide-based antibiotics which are
CC	active against a variety of organisms, e.g., bacteria, including
CC	multi-drug resistant pneumococci and other respiratory pathogens, as well
CC	as viral parasitic pathogens, or as crop protection agents (e.g.,
CC	fungicides or insecticides) via expression of polyketides in plants.
CC	Sequences AAZ87295-28302 represent macrolide biosynthetic genes from
CC	<i>Streptomyces venezuelae</i> ATCC 15439, which encode proteins
CC	AAV77190-Y771197.
XX	
XX	
Sequence	11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;

Query Match 41.4%; Score 41.8; DB 21; Length 11220;
Best Local Similarity 63.4%; Pred. No. 0.0041; Mismatches 37; Indels 0; Gaps 0;
Matches 64; Conservative 0;

QY 1 CGTGGCTGGGCTGCTGCTTCCTCGACCGGCGCTGATCCTGGGATGGGTTCAA 60
D 4362 CCGGCGCGGTGCTCTCAACACCCCTCCCGGAGCGGCGGCGGCTTCG 4421
QY 61 TGAGCAGGGCCCTGACTCGTTGATGCGGCTGAGATCCGCA 101
D 4422 TGACCTCGGATCGACTCGCTGACGCGGCTGAGCTCCGCA 4462

RESULT 11
AA28715
ID AA28715 standard; DNA: 15872 BP.
AC AA28715;
XX
DT 01-SRP-1997 (first entry)
DE Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KM Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KM polyhydroxybutyrate; biodegradable polymer; vep gene;
KM metabolic engineering; ss.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 20..13912
FT FT /*tag= a
FT CDS 14056..14136
FT FT /*tag= b
FT CDS 14148..15827
FT FT /*tag= c
XX
PN WO9722711-A1.
XX
PD 26-JUN-1997.
XX
PF 18-DEC-1996; 96WO-US20119.
XX
PR 19-DEC-1995; 95US-0008847.
XX
PA (MIND) UNIV MINNESOTA.
XX
PI Sherman DH, Williams MD, Xue Y;
XX
DR WPI, 1997-341701/31.
XX
DR P-PSDB; AAW19629-30 AND AAW00918.
XX
PT Expression cassettes for production of polyhydroxyalkanoate(s) -
PT provide wide range of biodegradable polymers for medical or
PT industrial use
XX
PS Claim 54; Fig 23; 91pp; English.
XX
CC Streptomyces venezuelae vep ORF1 (AA28715) comprises the polyketide
CC synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC also AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5'
CC loading module and a 3' end domain. Each of the sequenced modules
CC includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC reductase and an acyl carrier protein domain. The gene cluster was
CC cloned using a heterologous hybridisation strategy from a genomic
CC DNA library. A novel expression cassette encoding the first module
CC from the vep gene cluster and module 7 from the Streptomyces tyli
CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC activity and can be used for PHA prodn. in host (esp. insect) cells
CC for use as a biodegradable polymer.
XX
SQ Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

Query Match 41.4%; Score 41.8; DB 18; Length 15872;
Best Local Similarity 63.4%; Pred. No. 0.0043; Mismatches 37; Indels 0; Gaps 0;
Matches 64; Conservative 0;

QY 1 CGTGGCTGGGCTGCTGCTTCCTCGACCGGCGCTGATCCTGGGATGGGTTCAA 60
D 2833 CCGGCGCGGTGCTCTCAACACCCCTCCCGGAGCGGCGGCGGCTTCG 2892
QY 61 TGAGCAGGGCCCTGACTCGTTGATGCGGCTGAGATCCGCA 101
D 2893 GGAGCTGGGCTGACTCGCTCATGCTCGAGCTCCGCA 2933

RESULT 12
AA287283
ID AA287283 standard; DNA: 15872 BP.
AC AA287283;
XX
DT 05-JUN-2000 (first entry)
DE S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
XX
KM Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM chronic obstructive pulmonary disease; respiratory inflammation;
KM hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
FH Key Location/Qualifiers
FT CDS 20..13912
FT FT /*tag= a
FT FT /product= "vep ORF 1 amino acid sequence #1 (AA28717)"
FT CDS 14056..14151
FT FT /*tag= b
FT FT /product= "vep ORF 1 amino acid sequence #3 (AA28719)"
FT CDS 14167..15827
FT FT /*tag= c
FT FT /product= "vep ORF 1 amino acid sequence #2 (AA287178)"
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 96WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MIND) UNIV MINNESOTA.
XX
PI Sherman DH, Lau H, Xue Y, Zhao L;
XX
DR WPI, 2000-160679/14.
XX
DR P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methymycin and pikromycin -
XX
PS Example 3; Figure 23; 438pp; English.
XX
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eyc gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the

KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM chronic obstructive pulmonary disease; respiratory inflammation;
KM hypercholesterolemia; crop protection agent; ds.
OS Streptomyces venezuelae ATCC15439.
XX
XX WO20000620-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US14398.
XX
XX 26-JUN-1998; 98US-0105537.
XX
XX (MINU) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX
XX WPI; 2000-160679/14.
DR P-PSDB; AAY77180.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methylmycin and pikromycin.
PS Claim 13; Page 299-315; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC Streptomyces antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methylmycin,
CC pikromycin, neomethymycin, nadomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methylmycin, pikromycin, neomethymycin and
CC nadomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer syntheses or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439.
XX
XX Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;
SQ
Query Match 41.4%; Score 41.8; DB 21; Length 37948;
Best Local Similarity 63.4%; Pred. No. 0.0047;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 CGTGCCTGGGTCGTGGGCTTCCTGACCGGAGCGGCTGCTCGGATGGGTTCA 60
DB 21219 CCGTGGCGGTGCTCAACACCCCTCCCGAGGCGGTGACACAGGGGCGGCTTCG 21278
QY 61 TGAGCAGGCGCTGACTGCTGATGGCGGTGAGATCCGCA 101
DB 21279 TGACCTCGATTGACTGCTGACGCGGTGAGCTCCGCA 21319

RESULT 15
AAA75633 standard; DNA; 38506 BP.
XX
AC AAA75633;

XX
XX 22-JAN-2001 (first entry)
DF
XX Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
DE
XX
XX Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KM antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW pikromycin biosynthesis; ss.
XX
XX Streptomyces venezuelae.
OS
XX US6117659-A.
XX
XX 12-SEP-2000.
XX
XX 27-MAY-1999; 99US-0320878.
XX
XX 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
PI WPI; 2000-610844/58.
XX
XX
XX New recombinant pick hydroxylase gene of *Streptomyces venezuelae* useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
XX
XX
XX Disclosure; Columns 15-32; 117pp; English.
XX
XX The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in pikromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
XX Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;
SQ
Query Match 41.4%; Score 41.8; DB 21; Length 38506;
Best Local Similarity 63.4%; Pred. No. 0.0047;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 CGTGCCTGGGTCGTGGGCTTCCTGACCGGAGCGGCTGCTCGGATGGGTTCA 60
DB 18191 CCGTGGCGGTGCTCAACACCCCTCCCGAGGCGGTGACACAGGGGCGGCTTCG 18250
QY 61 TGAGCAGGCGCTGACTGCTGATGGCGGTGAGATCCGCA 101
DB 18251 TGACCTCGATTGACTGCTGACGCGGTGAGCTCCGCA 18291

Search completed: November 5, 2002, 13:10:44
Job time : 264.215 secs

SUMMARIES

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

6406.881 Million cell updates/sec

Sequence: 1 cgtcgtcgtgtgtctggct.....gatggcgttgagatccgca 101

Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 4109280

Post-processing: Minimum Match 0%

Maximum Match 1008
Listing first 45 summaries

Database :

```

1:  genbmb1: *
2:  gb_ba: *
3:  gb_hig: *
4:  gb_in: *
5:  gb_om: *
6:  gb_ov: *
7:  gb_pat: *
8:  gb_pl: *
9:  gb_pr: *
10: gb_ro: *
11: gb_srs: *
12: gb_sy: *
13: gb_un: *
14: gb_v1: *
15: en_ba: *
16: en_fm: *
17: en_hum: *
18: en_in: *
19: en_mu: *
20: en_om: *
21: en_or: *
22: en_ov: *
23: en_pst: *
24: en_ph: *
25: en_pl: *
26: en_ro: *
27: en_srs: *
28: en_un: *
29: en_v1: *
30: en_hig_hum: *
31: en_hig_iny: *
32: en_hig_other: *
33: en_hig_mus: *
34: en_hig_pln: *
35: en_hig_rod: *
36: en_hig_mam: *
37: en_hig_vrt: *
38: en_sy: *
39: en_hgo_hum: *
40: en_hgo_mus: *
41: en_hgo_other: *

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF217189	AF217189 Sorangium
2	101	100.0	68750	1	AF210843	AF210843 Sorangium
3	101	100.0	68750	6	AR193029	AR193029 Sequence
4	101	100.0	68750	6	AR199551	AR199551 Sequence
5	101	100.0	68750	6	AR199559	AR199559 Sequence
6	101	100.0	68750	6	AR199567	AR199567 Sequence
7	101	100.0	68750	6	AR201097	AR201097 Sequence
8	101	100.0	68750	6	AR208671	AR208671 Sequence
9	101	100.0	71989	6	AR172664	AR172664 Sequence
10	65.8	65.1	10910	6	AX024383	AX024383 Sequence
11	65.8	65.1	10910	6	AX024326	AX024326 Sequence
12	48.8	48.3	43280	1	SEU78289	U78289 Streptomyces
13	47.2	46.7	33529	6	AR166425	AR166425 Sequence
14	45.6	45.1	24225	1	SC2C4	AL512902 Streptomyces
15	45.6	45.1	65140	6	AX211705	AX211705 Sequence
16	45.6	45.1	104326	1	AB070940	AB070940 Streptomyces
17	45.6	45.1	123580	1	AF2633912	AF2633912 Streptomyces
18	45.6	45.1	125401	6	AX211739	AX211739 Sequence
19	45	44.6	28732	1	AF082100	AF082100 Streptomyces
20	44	43.6	32870	1	AF007101	AF007101 Streptomyces
21	43.4	43.0	15970	1	AE007061	AE007061 Mycobacter
22	43.4	43.0	53450	1	MTVO18	AL201889 Mycobacter
23	43.4	43.0	107379	1	SHGCP1R	X86780 S. hygroscop
24	43	42.6	20394	1	SNA132222	AX057922 Streptomyces
25	43	42.6	20394	6	AX057956	AX057956 Sequence
26	43	42.6	84985	6	SMA278573	AF278573 Streptomyces
27	42.8	42.4	49736	1	AF319998	AF319998 Stigmatel
28	42.4	42.0	5676	6	AE9718	AE9718 Sequence 1
29	42.4	42.0	53784	1	AMM223012	AJ223012 Amycolata
30	42.4	42.0	53789	6	AE9720	AE9720 Sequence 3
31	42.4	42.0	90445	1	AE040570	AE040570 Amycolata
32	42	41.6	82746	1	AE433501	AE433501 Actinosynn
33	41.8	41.4	30000	6	AX250261	AX250261 Sequence
34	41.8	41.4	30000	6	AX250262	AX250262 Sequence
35	41.8	41.4	37948	1	AF079138	AF079138 Streptomyces
36	41.8	41.4	78210	1	AB070949	AB070949 Streptomyces
37	41.8	41.4	107379	1	SHGCP1R	X86780 S. hygroscop
38	41.4	41.0	9513	6	AX089460	AX089460 Sequence
39	41.4	41.0	36181	1	U00023	U00023 Mycobacteri
40	41.4	41.0	41782	1	AY118081	AY118081 Streptomyces
41	41.4	41.0	50000	6	AX089417	AX089417 Sequence
42	41.4	41.0	50000	6	AX089420	AX089420 Sequence
43	41.4	41.0	80161	1	AY007564	AY007564 Saccharop
44	41.4	41.0	80161	6	AR155018	AR155018 Sequence
45	41.4	41.0	308950	1	MLERT9N	AL583925 Mycobacte

ALIGNMENTS

RESULT 1					
AF217189					
LOCUS	AF217189	58733 bp	DNA	linear	BCT 09-JUN-2000
DEFINITION	Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster; complete sequence; putative membrane protein gene, complete cds.				
ACCESSION	AF217189				
VERSION	AF217189.1	GI:7453554			
KEYWORDS					
SOURCE	Polyangium cellulosum.				
ORGANISM	Polyangium cellulosum Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Mycococcales; Sorangineae; Polyangiaceae; Polyangium. 1 (bases 1 to 58733)				
REFERENCE					

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-643 (2000)
MEDLINE 20115995
PubMed 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 10831849
PubMed 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submission
Submitted (16-DEC-1999) Korean Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source
1. 58733
/organism="Polysangium cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
<1. 992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62378.1"
/db_xref="GI:7453535"
/translation="VRG/FAFADGCGVPLLYDNLKNAVVRPHGDAIRPHITLAL
DSVYREPRPVAVARCNKCGVPRARVYRSGPRRAVADICDNRKQTEMTSSAL
SDSWEDPAPVYQADPDVRSVLRPPDPFPHHEVVEVCGTTPARNDKSTSPH
DTRRTLYLADISQRIQDGNIVATHRSDRGQIDQEPHQIDLVKDKRASHR
GIDRLAKNAKSSQAFIRTIKRGDNGSIAIKLQIDLVGADEKALVEVERDTI
HGAVAQYIDRRSERHLPVPSIFPKGEHALVYTPHSLTYIDALKDPTP"
999. 1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62379.1"
/db_xref="GI:7453535"
/translation="MTDLPTFKDKLKSIGLGLACMEQLADKFWLREVLAIERE
RHRSLERLKNRVAARFPMTPDSSMPKSIDREAVDOLYDSRADLLEFVTRTYD
AQRPLSLTNKAFADWGVFPAACVYTLVDRVHRAVEIIEAESYRKEKELNAT
RTKORTKKH"
1998. 6263
/gene="epoc"
1998. 6263
/gene="epoc"
/note="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62830.1"
/db_xref="GI:7453557"
/translation="MADRPFERAEDPIAIVAGCRHPGCVTDLSGFWTLLGSRDVT
GOVPAERMDAAAFDPDDIGKTFPTNRASPLSDVAFSGFISPRALMDGDPARH
LLEVCWELNENALAPSAIVGTETGVFTIGIPSEAEALPRTAETAEILADGGLGT
PSVAGRISYVLGLRGCYAVDPAVSSSLVAHVLGCRSECSALDAGVSLMLSP
STLVMLSKTRALATDGRCKAFSAEADGFGFCACVAVIKLSGAAADDDRLAIVRG
SAIINDGASSGLTPNGSSKEITVLKALADSCAASVGVYLAHGQTLLGPIETQIA
LNAVYGLGRDVPPLIGSKYKTNLGRPEVASITGLLVYLSLQHQDIPAHLLAAALN
PRIISGDLTLVTRARTPMVDMNTPRAVSSFGSCGTAHVAVLEAPAACTPPAPE
RPAELVLISARTAAALDAHARLRDLHLETVPSQCLGDVAFSLATRSAREHLEAAAT
RSEGLAALDAAAGCTPPGCVVAGIADSSRGKLAFLFTGGAQTIGMGLADVAF
RSEGLCVRLFNQGLDRPLIYEMMAEPASVDALADQATFOTPALFTFEYLAALARS
MGVEPELVAGHSIGELVACVAGVSIIEVAVFLVARGLMALPAGGAMVSIAPFA
DYAAAVVAPHAASVIAAVNGPDVVYIAGAGCPVHAIAAAMARAGARTKALHVSAPFA
PLMAPLEAFGRVAVSYHRSIVLVSNLSCAGDEVSFGYVWRHAREVYRFGLW
VKALHAAGAGTVEVGRKSTLLGLVPACLPDAPRALLLASSRAGRDEPAPVLEALGLM

AVGGLVSNAGLFPSSGRRVPLPTTYPMORERWIDTKADDAAGDRRAGAGDEVEKG
GAVRGDRSARLDHPPPSGRRKVEAAGDRPFLIEDPCVDRLRLVTRERAPG
LGEVILADNAGLSTFNQOLALGVPDDLPGRNPRLIGGECGDRYAVAGVAGVAG
VCGOPIATASAGAFHVTTSALVLPDQALSATPAAMAPAVLYLAWALGIALGV
GERVLIHATGCVGLAAYONAHQWCAVHATRGKERRAYIESIGVRYVSRSRDFV
ADPAAVYGESVDVYVNSLGGELDKSRNLKSRFRVLDGKDDYAAQNGSLREFR
NDSFSLVDKGMMLERPARVRLLELLAGVFTPTPTTLFIAVADHRSMAO
AOLDKRLVTDDEPVOIRIPIHAGAGSTGBRDLKLASAPARAALEAFRLTO
VSQVIRTEIRIGARVLFRLQDMLKVELNRLIENKLSKLSLFTSLSTPNIALTI
QMLDALATLSLEVAAMENLNAVQSDVSSGADQDMETIAL"
6260. 10492
/gene="epoc"
6260. 10492
/gene="epoc"
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINOLNELHOGVKLAADGERLQIQAPKNAIPLNLTARISEH
KSTILTMLRORLPASIVAPAPARHVPPLNDIGQSWLGRGATVPSGIHAYREYD
CTDLVARSRAFRKVVARHMDLRHATLPEDMQVTEPRVADIEIIDLGLDRSTREA
RLVSLRDAMSRIYDTERPILYHVAVRLDEQOQRLVSLIDLINVLGSLSTIFKDWL
SFYEDPETSPLVLESYNDYVLALESRRKSEHORMDYMKRRAVELPPMLPKAD
PSTLRIRPRHTEOMLPSDSMSRLKQVGERGILPTGYVILAAFSVIGRMSAPETL
NTTLRNRLPVPHPRVNDITGDFTSMVLIDITIDTRDKSFEORAKRIOEOLWEMHCDVS
GTEVORBARVIGLORGALFPVLTSLANQOYGVTSIGRGTTPYVSTQTPOLLDDH
OLYEHDCGLVIAEMDLYDCVFPDDILDDMLAEAVATLATEPMSORCSLPAOLE
ARASANENRSLSEPLTIGLIFAPARJOLPMQALVSAKRTLTETEISRRSRLCARIV
BOGAPNLTAAVAMERKGOYAVLVALESQAAPMDIADPAEIVHLLDHGKVIY
LQPHLIDKLSMPFCIORLVLDACVDEKQDQAPMPTIDSPADIAVYVITGSGSLBK
GVNIDHGRGVNTIIDINERTEIGDPDRLALSLSEFDSLVDVEIGILAGGTLYVPA
SKULPQAHMAALIEREKYVYMASVYALMLKDLVLEHSGEGLASRLSLKDSGDTIVP
GLPELOAIRPGVYISLGGATEASIKSTIGFVRVADVSMSIPIGRRLKQGTIVLD
GALIEPRPWWPQGLTIGGVLALGTWRDEKTRNSFLVHPETGERNLITGDGLRIPLD
GNIEFMRGENDQIKLNGTRVELGELEETLKLSPKPNDAVIVPVQDANKMLLAVVP
EGTRRRAEODDSLATERVDAAHAAKADKDEGGRVQFKLARHLDGKRPVLD
TCLVPRGELDVARRRSVRTLEAPIFPVEGRRLSLCSLSVEPDGALPFRKPSAG
STYPQVOTYAAKSGRIEGVDEGEFYVYHPEHRLKLVSHGLERGHAPQNDVDEQA
FGLFVEGRIDALIESLYGSLSEFCLLEGVAAQMLMEQAPSCNIGCVGQGFDEQVR
PLVLRDHSQVYVHGMAGRVDRQOVCTIGQDSSPRATRGAPGRDQHRADILRD
FLVRLPEYMPVTEVELDALFTSNGVYDRKALKERDTSRPSRSGTAPRDALIEI
LYAAVREVLGLEVVOLOQSFVDLGATSIHYAMRLLQKRIDREIATITELROYPLGS
LASGLRDSKDIQEQPNMODRVEARRKRRRS"
10639. 16137
/gene="epoc"
10639. 16137
/gene="epoc"
/note="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEQDSSAIVNIGSGRRPPCARDDDERVNRNDCTEAVQPSQ
ELASGVDPALVLDNRYHAGSVLDEVDPRFDAAFGISPRELEDDPDRITPMEAW
ALENGYDTPATGESIGYVAGCANMSYSLTSLNHEPRAKMLPAGTQTLIGNDKQLAT
HVSGLTGLPSTISVQTKSTLVANHLACSLDREDCDMLAGITVIRTRKQVTV
AEGSIFSPDGRHARDAKANGITWKGCGVILKRLDGLDGDGVLESLATVSNVND
GARKIGTFPDSHVEGQAQALMEALAGVAVASIOLEIHNHGTGTLGDALETRALR
DDASTYRSCALGSKVKTIGHLESAGIAGLITVLALEHQPLPSLNEESPNSIDF
LASSPYRVNTSLDKDWTGSTPRRAGVSFGIGTANNAVLEAPAAKILPAAAPARSAL
FVNSKSAALDAAARLRDLQAHQGLSLGDVAFSLATRSAPMHRILAAKAPVSEAL
REGDAAARGQTPGAVVRGRCSPGVNPRVVEFPGQSQWVGMRQLAEDPVRHVAL
SACDRAIIEAGWMSLALAEADGSSQIERLDIVVQPVYFALAIAVFAALMRSGVQPVY
VIGHSMGEVAALHVGALSLDEAVAIICRRSLILRIISGCEMAVTEILSLGAEALR
GLEDVSAVNSPSTVLSGEPALIGVLSLNAKGVCFRRKVDVSHSPQVDFLR
EDLLAALGILRRRAAVPMRSTVTGAMAGPELGNVMMNLRDQVRFAPVVOALQG
GHGLVEVNSPHILTTVSEEMRAORAGAAVGSIRRODDEPAILFALGLAMAGYV
VPMGRLPFAGGRRVPLPTYPMQREERYIEAPKSAAGDRGVRAGCHLLGEMQTLST

QSTRLEMTTLDLRLPLWLDGRHVQAVFPGAALLEMAISSGAELDGLQITDVA
LAEALFAGDAVAVLYOYVTEQPSGRLOFOLASRPGSHSPRHAGGLARVERE
VPAGTILSAVARILIOASMPAAATYAEITTEMLOQPRAGOGIAELHMGCEALGRRLP
DAAGSAEYRLHPLALDNCFOYVSGSLRAGGEBATPWVEVGSLELDREPGSELCHA
RVNHRQTPDROGADFWYVDSGAVAVAQVLAQRLPGRVRRREEDWLELEMER
AAVGTAKVNAQRMLLGGGGIGALRSMELAGHAAVHAESNTSAAGVALLAKAF
DGOPTAVVHLGSDGGELDPGAGCALDAPSPADVSPALPDRGCDGVLMVY
OALAMGFRDAPRLMLTRGONAVGAGVSTOAPLIGISRTIMEHNDLRCARDLD
PTPRDGEGLALLAEILDADAEEVALRGECVARIYAROPETPRKGTESCVPTDVT
IRADSTIYVTEGGLGSLVAGMLERKAGHLVIVGRGAASVQRAVAALLENAGAR
VTAKADVADRAQLERLIREVTTSGMPLRGVHAAGLIDDLMOQTPARKVWAPK
VOGALHHLALTREAPLSEFVYLVAGVGLSPGGONYAANTFLDALHHRHAGLPV
LSVMDLFEAEVMAAOEDRGARLVNSGRMSLTPEDEGLALATLLESGRVOGVLPV
PLWVEIYPAASRMLSRITVATRASAGSPAGBDILRLIAAEPSSRSGLEPILR
AOISQVLRLEPGKREVDARPLTSLGMSNLMSGLELNRIRIEMGLITVPTLVVAA
LSGHLAREACAAAEVESPHTTADSAVEITEMSODDLQOLIAKFRALT"

gene
16134..37907
/gene="epod"
16134..37907
/gene="epod"
/note="type 1 polyketide synthase; epothillone PKS modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=1
/product="epod"
/protein_id="AA56283.1"
/db_xref="GI:7453560"
/translation="MTTGTPTAQNPLKQAIIITQREERLAGLAQAELEPTETAIY
GIGRFGGADAPAEFWELLDAERDAVQDRLRALVAVAEVPIPHAGILTEPIDC
FDAEFGISPEARSIDPOHRLLEVAEGLEDAGIIPRSIDSGITGVFAATFADYA
RTVARLPREEDASATGNMLSIAAGRLSTYGLGOGPCTIVDTAKSSSLVAIHACRS
LRAGESDIALAGVSTLSPDMEAAARTOALSPDRCTEDASANGVREGGCGVY
LKRISDAQRDGRITMALIRGSAINHDSRTGLRPANVLAQETVIREALRSAREGAN
DYVTEHGRTSLDPIEVEALRAVGPARKSDGTQRCVJGAVKTNIGHLEAAAGVAGLIK
AALSLTHERIPRNLFRLNRIELRGSALATPEVPMPTDRRREGVSGFSGST
NAHVLEAPAVEILMPAPAPERSAELVLSGSEGALDQAARLEHLMHDELIGDY
ASFSLATRSAMTHRLAAVATSRGELIALVSAOGOPAGAAARLASSRGLKLLFT
GOGAOTPGMGRLCAAMPAREARCTYLFREDRLREVMMAEAGSALLDPT
AFQPALFAVEYATATLMRMSGVPEELLVGHSTIGSELVACVAGFSLDEGRVLAAR
RLMGSLSGAMVSLGAPAEVAAVAPHAWVSIIAANGPEQVIAQVEADAVOATIA
GFAARGVTRKLIHSHAFHSPMLMEPMEEFGRVAAVYRRPSVLSVNSLKGAVTDE

Query Match
Best Local Similarity 100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CGTCGCTGGGCTGCTGGGCTTCCTGCAGCCGCGCTGCATCCTGGAGGAGGTTCAA 60
|||||
Db 24950 CGTCGCTGGGCTGCTGGGCTTCCTGCAGCCGCGCTGCATCCTGGAGGAGGTTCAA 25009
|||||

Oy 61 TGAGAGAGGCGCTCGACTGCTTGATGGCGGTGAGATCCGCA 101
|||||
Db 25010 TGAGAGAGGCGCTCGACTGCTTGATGGCGGTGAGATCCGCA 25050
|||||

RESULT 2
AF210843 68750 bp DNA linear HCT 21-Nov-2000
DEFINITION Sorangium cellulosum strain So ce90 epothillone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE
ORGANISM
Polyangium cellulosum.
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangiineae; Polyangiaceae; Polyangium.
1 (bases 1 to 68750)
REFERENCE
AUTHORS
Molnar, I., Schupp, T., Ono, M., Zitzke, R., Milmanow, M.,
Nowak-Thompson, B., Engel, N., Toupet, C., Straumann, A., Cyr, D. D.,
Gottlieb, U., Mayo, J. M., Hu, A., Goff, S., Schmid, J. and Ligon, J. M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothillones A and B from Sorangium cellulosum So ce90

JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar, I.
TITLE Direct Submission
JOURNAL Submitted (03-Dec-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd., P.O. Box
12257, Research Triangle Park, NC 27709, USA

FEATURES
source
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=1
/product="unknown"
/protein_id="AA56290.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSARVLDGGRATERVHVAEARGIEDLRALREH
LRIOEGSPFCMCIGDITVELLADHDPLASISFHHARSRLHPDWTSDAMLVDPALV
RMLARCAPGLREYEERERARTAOERLMLAAAPCPADLPREFEDANGLP
MSPEVAERLRRLASATPELACALAILAMLTGSGPMGSPAYEMERLPTNLGLPLP
AIAASAPGTESEALRGARLPASVEVSSKSGOLGPEALMERLPTIVRAMGNAD
LSREBRALIAEYRLRAQAPAPRAGAGLVAGVSSSGRLSGVLTGDDALYSGGND
IVMFQGRISIPVLLAGTDPFELAPPLSQMLFAVHANAAGTISKVLEGSPLIVARN
QARPSMLHWAGFAMVWNOQAVPDERCAPVYORSTIMEFHPTRCLHPASAFS
LACEHEHYMCESLGRLELRRHPRGASPRAYLGEHPILATWYPSLTNLNTHVL
WADDDRRILVQVDRGVEPVLAETRHRPAPVYSEDIDIALGQDPSRDMVHEHR
SGASVYVADYORLMDRPDMINRGLEPFTINDILTIARS"
complement(1900..3171)
/note="ORF2; similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to DD-peptidases"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AA56290.1"
/db_xref="GI:6724239"
/translation="MGALISVAAPGALGAEERGQPDGAGALAPAREMAEVA
AGMDGVAWLVARGDVDVDAVGYTELGSAFMRDRTFRIASMTKATVAAVMLE
EGKIDISPVDRMPELANRKLVARIOSPEIDYPAERPTVRMLPTMGFISFDA
SSPTORATIDELGVNAQVPMTPHGPDEMTRIGLTPLMHOPGAMYNTGSLVOGYL
VGRADQGEFDAFVERRLIAPLGMDDTDHVPADKLARAGCGYFTDEGTGKTRNDRO
GESAAYASPAPFPGSAGLVSTVDYLLFAFAMLNKGVHEGRRLISAAYEMTADHL
TPAQKAASFPGFEETHGMGYMAVYAPAVSEVPCRYGMDGSGFTSWINDGREL
IGIYWTGASGFLFGALERFRSYVAATESA"
3415..5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AA56290.1"
/db_xref="GI:6724240"
/translation="MHGTERQVILSTVTLALIVTARASELIARLRQPYDELRG
GGVILGSPVAGALAPCFHRLAEQPAVGVLSIGTWIGLILLMAGLEVDDGILRKEA
RPGALSALGAIAPLPAAGASALVLDPLGGLGVLSLSTAVYAKVLIBREMS
RRSYAOVTLAGVYSEVAAVILVAMTSSYCASPALAARASALLAGFLVLMVAPRT
LTHLAMRVNADATRSKGOVSLVILVFLAALORGLPHLGLHAFGLVLLNAPR
MRPLIDGYQTVIAGLFAVFFVLAMGRVDSQLTRPAMGVALLATATKAVVPA
IGARLGLRGSEBALVAGLMMKGCTDITVAIVGEGLISNEATYATVAAVTVTA
SPALLIWLKRAPPEQESARLEEEAARAVIPQVERIIVP IVAHALPGATVIVES
IVASRKLGEDVIDITELSEDOAGPSPAGAEASGLARLARLVGIVWRRLRGS
IOALIRASRDHDLVITGARSPARAGMSGLQDAIYORASNVLVVYGDPPAERAS
ARLIVPTITIGEYSPAADLAHAHLVADALVLSAQDTPGAVVMDRPPRAV
ARSYVDAVEFRGRUGVSSRVHVGAHSPSEITRELARAYDILVLCYDHGLGRLL
YLGSTVESVYRSRYPVALVYAHGSTRQVR"

CDS
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=1

/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGRTRCAIAEGYIPGSGNGPEPOMTSHECTACLLNADSDRA
OHTATVYFSDRDPACGYPRTVPAARPTRHFRPNLDEPEPIPRDTDAASYIESAPLVV
OHTLRDSROAENALLSTIAVTBRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724342"
/translation="MKHIDTGRFRGRRIGTLLGLASMLAMAGCGSEKTVGSTRAP
GADARVATADVDPDAATTLIAVDVHLSPERLEAGSEREPVMPORPSPESPWRVGLD
VYADSRGKLAETVYVYANPELLITAEKOSSPGSSAAVIGTYSV"
7610..11875
/gene="epoa"
7610..11875
/gene="epoa"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADPEIERAAEDPIAIVGASCRLLPGGVIDLSGFWLLSESRDT
GVYPARMDAAAMPDDPDPAQCKPTVTRASPGLSDVACPDASFGCTSPRRALMDPAR
LLENCWEALENAIATPNSLVGTETGVTGIGPSEYEAALPQATASAEIDAHGLOTM
PSYAGGRISTALGRPCVAVDTAISSTLVAAHLAQSLRSGECSTALAGVSLMLSP
STLVLSKTRALARIDGRCAFSAEADGEGRGECAVVLRLSGARADADRIILAVRG
SAINHDGASSGLTVPLIGSOEIVLKRALDAGCAASVGVYEAHGTTLLGDEITQIA
LNAVVGIGRDVATPLIGSVKTNLGHPEYASGITGLKLVLSLHGQIIPAHLLAQILN
PRISMGDLRTVTRARTPMDWTPRRAGVSSFGSGTNAHVLEAPATCTPPAPPE
RAEILVLSARTASALDADAARLRLDLETYPSCGLDVAFLATTSAAENHRLAVAT
SREGLEAALDAAAGQOTS:GAVRSIADSSRGKLAFLFTGOGAOTLGMGVLVDWMAF
REAFDLCVRLFNOELDRP:REVMMAEPASVDAALILDOTAFTOPALFTPPVLAALMRS
MGVEPELVAGSTIGELVAVCAVGSFLEDAVFLVARGLMALPAGCMALVSEAEPA
DYAAVAAPHAASYSIAVIAVNPQVYTAGCOPPHATIAAAMAAGATKRLVHSHAFS
PLMAPLEAFGRVAVESVTRPSIVLVSNLGACCTIDVSSFEIYVRAHREYVRADG
VRLHLAGAGTEVEGPK:TLGLVLPACPKDAPALDASSRGRDRPAVLELGLGM
AVGLVSMAGTEPSSGGRVPLPTVPMQRRVYDITADADAARADAPAGHDVEEG
GAVERGDRRSARLDPPPI:SGREKVEAAGDRFRLIDEPEVLDRLVIRVERRAG
LGEVEIADVADAGLSFNDVCLALGMVDDDPGRKPNLLIDGEGAGIVAGCEVNGLV
VGQPVIALSAGAFATHTVTSALVLRPGALSAIEAAMPVLEIATMYALDRARLOP
GSRVLIHAATGGVGLAAVCAQHVGALEVATAGTPKRAYLEIAGRYVSDSRDEFV
ADVRAWTGEGVDVYVNSL:SGELIDKSFNLHSGRFVEELGRDCYADNOLGRPLR
NLSEFSLVDLRGMMLERPAVRLALLELLGLIAGVTPPIPTIPARVADAFRSMWO
ACHICKVLTLLDPEVOIHIPTTHAGGSPAGTDRDLDRLASAPARAALAEFLPTO
VSQVLEPRTIKVGAELPTIRICMDSIMAVELRIRITASLKIKTSTTFTSTSPRIALLA
QNLIDALATLSIERVAALNLRAGVONDEVSSGADQDMETIAL"
7640..8920
/gene="epoa"
/note="EPOS A"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
/gene="epop"
11872..16104
/gene="epop"
/note="EPOS P: contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11

/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINOLNELEHOGIKLAADDERLOIOAPKNALNPNLLARISEH
KSTILMLRQRPASISIVAPAREHAPPLNPIOGSMVGRGATGATVPSGIAHREYD
CNDIOPRLSARFRVYARHMDLRHATLPDMNOVLEPYVDADITETIDRGDRREA
RLVSLPDMASHRIYDTERPPLIYAVARLDEQTSLLSIDINWDLGSLSTIFKDWL
SYEDDESLPYLEISRDYVALLESRRKSEHQSMQMDWKRRIMLEPPLPKAD
PSTLEIRNRHEQMPSDSKGLRKRGERCLTPTGYTLAFAFSEVIGKWSAPFTL
NITLERLPHPRNDITGDETSMVLDDITDTRDSFEDRARIDQLMEADHCDVS
GIEVOREARVGLIOGALPEVLTLSALNOQVGYSLQRLTGPVYSTQEPQLLDH
QLEHGDVLIAMIDVGVFPEDLLDMM:AAVVLRLRTEPEMGEBOVRCSLPPQLE
ARASANATNALISEHTLHGLFAARVEQLEPMQAAVVSARKTLYEELSRSRGLARLR
EGARPNTLVAAYMEKMGEOVAVVLALESAGAAVYIDADLPAERLIHYLLDGEVRLV
LQOPMDKGLSMPGJQRLVLEAGVEGGDDPPMPLQTPDGLAVVITYTSSTGLPK
GYMIDHRAVNTLIDINERFEETGGPGRYALSSLFDSLVDYVDFGIIAAGGTIVPDA
SKLRDPAHNAELIEREKYVWNSVVALMMLVYHREGPRDSLARSRLRSLISGDMIPV
GLPGELOAIRBGVSYSISGAGTASINSGYPRVNDLSMASIPYGRPLRNTFFVLD
EADERPVWPCQLYTIGGYLALGTWRDEKTRKSFVHPETGERLKITDGLRTLPD
GNIERGRDNDQIKLNGYRVELEIETLKSHPNVDAVIVPGNDAAKLLALVLP
EGTRRRAEQDASLATERIDARAHAAEADGLSDGRVQKILARHGLRDLDGKPVVDL
TGQDPRKAGLDVYARRRSVTELEAPIPEVEHRL:SCLSVEPDPGATLPKRPYSAG
STYPVOTYAVYKSGRIEYDEGFYHYHPEHRLRLSDHGERGAHVRONPEDEAA
FNLFEVGRIDAIESLGSSSREFCLLEAGYAMQLMEQSCNIGCVGQFNEQVRA
PYDLRHSVDVYHMGIGSRVDPROVCTLGODSSPRRATGAPGRGROHPADALRI
PLTKLPEYMWTFVEIDALPLTSGVYKDRALREKDTSPRSRGTAPADALIEI
LVAAYREVIGLEEVGLQOSFVLDGATSIHIVRMSRLQKRLDRELAITELTPYPLGS
LASGLRDSRDLDRPNMDRVEVRKGRRS"
16251..21749
/gene="epob"
16251..21749
/gene="epob"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match 100.0% Score 101; DB 1; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.8e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGTCGCTCGGCTGCTGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATGGGTTCAA 60
Db 30562 CGTCGCTCGGCTGCTGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATGGGTTCAA 30621
QY 61 TGAGCAGGGCGCTGACCTGCTGATGGCGGTGGAGATCCGCA 101
Db 30622 TGAGCAGGGCGCTGACCTGCTGATGGCGGTGGAGATCCGCA 30662

RESULT 3
LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Medison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Goriach,J.
TITLE Genes for the biosynthesis of epothonones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES Location/Qualifiers
source 1..68750
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match	100.0%;	Score 101;	DB 6;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 1.8e-15;		
Matches 101;	Conservative	0;	Mismatches 0;	Indels 0;

Qy	1	CGTCGCTGGGTGCTGGGGCTTCTCGACCCGACGCGCTCGATCCTGGGATGGGTTCAA	60
Db	30562	CGTCGCTGGGTGCTGGGGCTTCTCGACCCGACGCGCTCGATCCTGGGATGGGTTCAA	30621

Qy	61	TGAGCAGGGCCCTCGACTCGTTGATGGCGGTGAGATCCGCA	101
Db	30622	TGAGCAGGGCCCTCGACTCGTTGATGGCGGTGAGATCCGCA	30662

RESULT 4	
ARI199551	
LOCUS	ARI199551
DEFINITION	Sequence 1 from patent US 6355457.
ACCESSION	ARI199551
VERSION	ARI199551.1
KEYWORDS	GI:20249625
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 68750)
	Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
	Gorlach,J.
	Genes for the biosynthesis of epothilones
	Patent: US 6355457-A 1 12-MAR-2002;
	Location/Qualifiers
	1..68750
	/organism="unknown"
BASE COUNT	9596 a 22456 c 25339 g 11159 t
ORIGIN	

Query Match	100.0%;	Score 101;	DB 6;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 1.8e-15;		
Matches 101;	Conservative	0;	Mismatches 0;	Indels 0;

Qy	1	CGTCGCTGGGTGCTGGGGCTTCTCGACCCGACGCGCTCGATCCTGGGATGGGTTCAA	60
Db	30562	CGTCGCTGGGTGCTGGGGCTTCTCGACCCGACGCGCTCGATCCTGGGATGGGTTCAA	30621

Qy	61	TGAGCAGGGCCCTCGACTCGTTGATGGCGGTGAGATCCGCA	101
Db	30622	TGAGCAGGGCCCTCGACTCGTTGATGGCGGTGAGATCCGCA	30662

RESULT 5	
ARI199559	
LOCUS	ARI199559
DEFINITION	Sequence 1 from patent US 6355458.
ACCESSION	ARI199559
VERSION	ARI199559.1
KEYWORDS	GI:20249633
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 68750)
	Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
	Gorlach,J.
	Genes for the biosynthesis of epothilones
	Patent: US 6355458-A 1 12-MAR-2002;
	Location/Qualifiers
	1..68750
	/organism="unknown"
BASE COUNT	9596 a 22456 c 25339 g 11159 t
ORIGIN	

Query Match	100.0%;	Score 101;	DB 6;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 1.8e-15;		
Matches 101;	Conservative	0;	Mismatches 0;	Indels 0;

Qy	1	CGTCGCTGGGTGCTGGGGCTTCTCGACCCGACGCGCTCGATCCTGGGATGGGTTCAA	60
----	---	---	----

Db	30562	CCTCGCTGGGTCGTGGGCTTCTCGACCCAGCCGCCTCGATCCTGGGATGGGTTCAA	30621
Qy	61	TGAGCAGGGCCTCGACTCGTTTGATGGCGGTGAGATCCGCA	101
Db	30622	TGAGCAGGGCCTCGACTCGTTTGATGGCGGTGAGATCCGCA	30662
RESULT 6			
LOCUS	ARI99567	68750 bp	DNA
DEFINITION	Sequence 1 from patent US 6355459.	Linear	PAT 20-APR-2002
ACCESSION	ARI99567		
VERSION	ARI99567.1	GI:20249641	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 68750) Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J. Genes for the biosynthesis of epothilones Patent: US 6355459-A 1 12-MAR-2002;		
TITLE	Location/Qualifiers		
JOURNAL	1..68750		
FEATURES	/organism="unknown"		
SOURCE	1..68750		
BASE COUNT	9596 a 22456 c 25539 g 11159 t		
ORIGIN			
Query Match	100.0%; Score 101; DB 6; Length 68750;		
Best Local Similarity	100.0%; Pred. NO.1.8e-15;		
Matches 101; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	CCTCGCTGGGTCGTGGGCTTCTCGACCCAGCCGCCTCGATCCTGGGATGGGTTCAA	60
Db	30562	CCTCGCTGGGTCGTGGGCTTCTCGACCCAGCCGCCTCGATCCTGGGATGGGTTCAA	30621
Qy	61	TGAGCAGGGCCTCGACTCGTTTGATGGCGGTGAGATCCGCA	101
Db	30622	TGAGCAGGGCCTCGACTCGTTTGATGGCGGTGAGATCCGCA	30662
RESULT 7			
LOCUS	AR201097	68750 bp	DNA
DEFINITION	Sequence 1 from patent US 6358719.	Linear	PAT 20-APR-2002
ACCESSION	AR201097		
VERSION	AR201097.1	GI:20251985	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 68750) Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J. Genes for the biosynthesis of epothilones Patent: US 6358719-A 1 19-MAR-2002;		
TITLE	Location/Qualifiers		
JOURNAL	1..68750		
FEATURES	/organism="unknown"		
SOURCE	1..68750		
BASE COUNT	9596 a 22456 c 25539 g 11159 t		
ORIGIN			
Query Match	100.0%; Score 101; DB 6; Length 68750;		
Best Local Similarity	100.0%; Pred. NO.1.8e-15;		
Matches 101; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	CCTCGCTGGGTCGTGGGCTTCTCGACCCAGCCGCCTCGATCCTGGGATGGGTTCAA	60
Db	30562	CCTCGCTGGGTCGTGGGCTTCTCGACCCAGCCGCCTCGATCCTGGGATGGGTTCAA	30621
Qy	61	TGAGCAGGGCCTCGACTCGTTTGATGGCGGTGAGATCCGCA	101
Db	30622	TGAGCAGGGCCTCGACTCGTTTGATGGCGGTGAGATCCGCA	30662

Db 30622 TGAGCAGGGCCTCGACTCGTTGATGCGGATGAGATCCGCA 30662

RESULT 8
LOCUS AR208671 68710 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US "383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
FEATURES
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source 1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred.No. 1.8e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGTCGTCGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATTCAA 60
Db 30562 CGTCGCTCGGTCGTCGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATTCAA 30621

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGCGGATGAGATCCGCA 101
Db 30622 TGAGCAGGGCCTCGACTCGTTGATGCGGATGAGATCCGCA 30662

RESULT 9
LOCUS AR172664 71985 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
FEATURES
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epoethilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred.No. 1.8e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGTCGTCGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATTCAA 60
Db 24950 CGTCGCTCGGTCGTCGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATTCAA 25009

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGCGGATGAGATCCGCA 101
Db 25010 TGAGCAGGGCCTCGACTCGTTGATGCGGATGAGATCCGCA 25050

RESULT 10
LOCUS AX024383 10910 bp DNA linear BCT 15-SEP-2000

DEFINITION Sequence 81 from patent DE19846493.
ACCESSION AX024383
VERSION AX024383.1 GI:10184587
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
REFERENCE 1 (bases 1 to 10910)
AUTHORS Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
JOURNAL Beyer,S. and Mueller,R.J.
PATENT: DE 19846493-A 13-APR-2000;
FEATURES
source 1..10910
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 1852 a 3124 c 3613 g 2321 t
ORIGIN

Query Match 65.1%; Score 65.8; DB 1; Length 10910;
Best Local Similarity 78.2%; Pred.No. 1.1e-06;
Matches 79; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGTCGTCGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATTCAA 60
Db 2000 CGTCGCTCGGTCGTCGCTTCCTCGACCGAGCGGCTCGATCCGATGGGATTCAA 2059

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGCGGATGAGATCCGCA 101
Db 2060 CGAGCAGGGGCTCGACTCGTTGATGCGGATGAGATCCGTA 2100

RESULT 12
LOCUS SFU78289 43280 bp DNA linear BCT 13-AUG-1997
DEFINITION Streptomyces fradiae tylactone synthase, starter module and modules
ACCESSION 1-7, (tyl) gene, complete cds.
U78289

```
VERSION      078289.1 GI:2317859
KEYWORDS
SOURCE       Streptomyces fradiae.
ORGANISM     Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
              Streptomyces fradiae.
REFERENCE    1 (bases 1 to 43280)
AUTHORS      Dehoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
TITLE        Sequence of streptomyces fradiae tylactone synthase gene tylG
JOURNAL      Unpublished (1996)
REFERENCE    2 (bases 1 to 43280)
AUTHORS      Dehoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
TITLE        Direct Submission
JOURNAL      Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
              Center, Indianapolis, IN 46285, USA

COMMENT
FEATURES
    source          location/Qualifiers
                    1..43280
                    /organism="Streptomyces fradiae"
                    /db_xref="taxon:1906"
                    1..43280
                    /gene="tylG"
                    816..14234
                    /note="ORF1; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="tylactone synthase starter module and modules 1
                    & 2"
                    /protein_id="MAB6504.1"
                    /translation="MSSLRRAVQSNCGYGDLMTSMTAONQNTDODVDGPDSTHGGE
                    IAVVMSCLRPGAGVEEFPFELLRSGRMPTRODDCTWRALIEDHAGPAGFGMNR
                    QAAATDPQRIHMLEGWEALEADAGTIPGDLITGDTVPFACVADSDYAVLTRRSVAG
                    GTATGLHRLALANRLSHFLGKPSLVDSASASLVAVQLACESLRKETSIAVAG
                    RYVNLITBESTVYMERMGALSPDGRCHTFDARANGYVRGSGGAVLKPDLADLADG
                    RYVCYIKGCAVNMDSGASLTLPDRAQAVLQOARRAGSVGAIVRYELHGTGRA
                    GDPVEAALGAVIGACAGDSRSTPLAVGSVKTNGVLEGAAGIVGLIKATLCYRKEL
                    VPSLSTPTNPVPIPLDLRLRNOTERQENEDDPRVAGVSFGMGNTVHIVIELA
                    PAAAGSGAGSGAGSAGACISAVSGVPPVVSRSRNVYREAARLAEVDEAGVGIA
                    DVAATMADSRGQYRAVVLARKEALEAGRLALAGDPAGVVTGAVLDGGVAVVAG
                    GGAGAGGAGAGAGAGGAVLVFPQGTQWGMGAGLGSSEVTAASMEKALVAP
                    HYGMDLLEVYSGAGLERVDYVQPTVMAVMSLARWQAMGVVAVAVGSGEILAA
                    TVAGALSTLEDAAVVALRAGLIGRYLAGGMAAVPLPAGVEARGLAKMPCGVAAVN
                    GCRASYVSGDRRAVAGYVAVCAQEGQARLIPDYASHSRHVEDLKGELETVLSGIRP
                    KSPRPVCSSTVAGEQGEPEVFDAGTWFRNLRNRYE SAVVGILLGEGHREFEVSAP
                    VLVALEQTAEARADSVHATGLRRQDDSPHRLITSTAEAMHAGATLTMDPALPGHL
                    TTLPTVPFNHHWMLDTIDGGGDDATOKESGPTLRRLGLPSOKQLGFLDLYCR
                    HTAAVVLGDLTAAVEPDLSEFKKQIQISMTGVELRNRLIETGLALPTTLTYDDPTRA
                    LAQFLHTELDGSPSGSVLAPAKSFEAOEPIAVVGMGCRPGCVGSPALIMRVYEG
                    VDAVSFPFCDRGDVEGLYDPEPGVAGKSYVRGGFLHDAEEDAFPRGIPREAVAM
                    DPQARLLLETSWALEIRAGIDPHSLHGSRTGVAGWPDQYGRLAGEGVSGYLLT
                    GTSGSVASGRVATLGLGPAVTVVDACSSILVALHVAOLGJECDMALAGVAVM
                    AGPGMEVEFSRQGLAADRCARAFADGADGTAAEAGVTLVERLSDARLGHPTVLV
                    VCGSAVNOGASNGLTAPSPGSOEYVIRIOLAGNARLTVDVDEVAHGTGTLGGDPLE
                    AQAIGTYGRDRGGRPVWLGLSKNIGHAQAAGVAYIKWILARNGYGLRPTLLDLE
                    ESRHAYDMSAGVWMLTEAREMPCVDPRRAVSAVSGCTNAHLLIEBAPTRAEASA
                    TTPRSEVSESAVVDARSQVVPVVVSGRSRVVRAAGLAVVPAVGGLADVAT
                    MAGSRFEGRAVVLARGAEALRLALAGGDDADVAGVADVPDPSGGGGLADVTF
                    PGQGTQWGMGAGLGSSEVTAASMEKALVAVHEMDLLEVYSGAGLERVDVQPT
                    VMAVAVMSLARWQAMGVVAVAVGSGEILAAVTVAGALSTLEDAAVVALRAGLIG
                    RYVNLITBESTVYMERMGALSPDGRCHTFDARANGYVRGSGGAVLKPDLADLADG
                    IGVAPARWALTRRAVAVVGPETPDAGAOIOWFVSLGLVPSDIADRTSTALLOQELT
                    TRTPETSOPQPTPERLPQTPNRRALFLAAVVLARCGEDOVAVRASGTYRRRSRAA
                    AGASWQPSGTYLITGGMGALGRRLARLARAIEAEVLVLSRGPAPGAELAEIRAK
                    GHCEVVAHACDVAERDALAALVTAPPAVAFETADILLDAVDITLSPSEFEIRAK
                    VCGAELLHQITADIKGIDAFVLFSSVTGWMNGOGAVYAAANALDLAELERRAALCP
                    ATSVAMGLWGGGMAAGGEESLSRGRGLRAMDPDAVVDALILGAMGRNDVCVTVDVDM
```

```
ERRAPATNMRPGRLEFVPEAREALTAAGTTSATPDGAPBELARLSMLNETERLKL
VELVETAEAFVLSHPNTDAIGAREPKSAGDSLTSLERLNRLNAGCTKPAVTLED
HSPRTALARLLDLRLTGACAPAPAADEPPLPVAADDDVPVITYGMACRPPGGAGTEA
LKRLVTEERDVITCAAPTRDKMDSDSYDEDEPGVAGTYYVRGEGFLHDAEEDAFEGCI
SPREAVAMPDQARLLETSWALEIRAGIDPHSLHGSRTGVAGWPDQYGRLAGEGVSGYLLT
EYEGVLTGKSASVSGRSISYTLGLGKPSLSDTPACSSILVALHNAQALRGECEDMA
LAGGVTYMAAPGLFEFSRQGLAADRCARAFADGADGTAAEAGVTLVERLSDARLGHPTVLV
LGHPTVAVVCGSAVNOGASNGLTAPSPGSOEYVIRIOLAGNARLTVDVDEVAHGTG
TRIGDPLEAQALLGTGRDRDAECPPWISGLSKNIGHAQAAGVAYIKWILARNGYGLRPTLLDLE
LRPTLVHDEPSRHWDSAGVRLITLTEREMPCVDPRRAVSAVSGCTNAHLLIEBAPTRAEASA
LEALEALDATDAPAPAPPAEDVTIDTEALEADTAEGAKARAFSGSEAPQVAVP
VVVSGRSRVVREARAGRLAEVDEAGVGLADVAVTAAHGSREGRYRAVVLARKEALEAG
RLALAGSDPDVAVAGVADVPETGSGGCVLVVFPQGTQWGMGAGLGSSEVTAASMEKALVAP
VMSRCEARLSTVHYEMDLLEVYSGAGLERVDYVQPTVMAVMSLARWQAMGVVAVAVGSGEILAA
VTVAGALSTLEDAAVVALRAGLIGRYLAGGMAAVPLPAGVEARGLAKMPCGVAAVN
KMPGVVAAVNGPASTVSGDRRAVAGYVAVCAQEGQARLIPDYASHSRHVEDLKGELETVLSGIRP
KSPRPVCSSTVAGEQGEPEVFDAGTWFRNLRNRYE SAVVGILLGEGHREFEVSAP
HLREFIEVSAPVLVALEQTAEARADSVHATGLRRQDDSPHRLITSTAEAMHAGATLTMDPALPGHL
TMDPALPGHLTTLPTVPFNHHWMLDTIDGGGDDATOKESGPTLRRLGLPSOKQLGFLDLYCR
DSGERVEAGRLAGEHMDLTDHVAHSVTLTGPAPVEFLAHAGATGCGLELSEVA
PIVLPAGCYRQMRVSADESGRRRVATLSAPEAAVHSAEGDSAGYTWTRGEGTL
VDPPEPTPDADBARAMPAGEVEVEBELTERGALTEYEGEAFASVRAVWRPDDL
AEVLLPDRASTAGRGVHPHALDAAIQPWIAGLLVEPDDAVLTFPAMOGVSLVAG
AGALRYRLTKAGDGAVSLQADTSGAAVSLIGALVWRPYARRLDVLTGDAGERSLY
RYEMQPRLLPAGPRSMAYLGPADRLAGCPGLGDQPDGPTALPEVRLARKALAG
APPEAVVLPVLSAGATPESVROTTERCLTALQDMLDEELVDPTLVILTRGAVAAV
PGEIEDIACAGWGLYRSASRHPGRFLVYDDHMDPDRPTALPLARVLQAGOLS
LRAGTRTPLYLLRAGTPEBORGPAFPDPAQVLYTGATGTLGRLLARHLAAHEVRHLL
LLSRGGRAGADELAELAELAEPCFAACDADAEALARVAIEVPADPLTGLVIA
AGVLGDGTLDALPERIGTWPRPKADANLHBLITPSPISVYAEVAGSAGLIGRGQ
ANYAAANTLEFDALAOHRAHAGLPAVSLIAMCAGCMGHSGLGTLRMRRSGLIAP
THOQGLAFPRALAAASDEPLLVPMRLDIAVLEAREAHGSDPDPPLGLILPARAY
ROAARVYRGAPAGGEGTARLGCATGICFARLRELVYRAASVCLYSGDDAEP
GRFPKLDGDSLTAVELRNRLGATGRLPTLPTAFPRPSQAVETLAELAEAPRPGG
DITAAAFEGIEALAAVAGALAEIDLRRVLRRLTELAALTYQGNPSPAPAPASDL
DERLIDANDDILPFIETEQI"
942..3929
/gene="tylG"
/function="encodes starter module"
942..2136
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase KS-Q of
starter module"
2571..3557
/gene="tylG"
/function="encodes acyltransferase domain of starter
module"
3675..3929
/gene="tylG"
/function="encodes acyl carrier protein domain of starter
module"
3993..8471
/gene="tylG"
/function="encodes module 1"
3993..5264
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase of module 1"
5631..6617
/gene="tylG"
/function="encodes acyltransferase domain of module 1"
7410..7949
/gene="tylG"
/function="encodes beta-ketoreductase domain of module 1"
8220..8471
/gene="tylG"
/function="encodes acyl carrier protein domain of module
1"
8541..13970
/gene="tylG"
/function="encodes module 2"
8541..9812
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase of module 2"
```


Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Blbb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or atc) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring subclones. Cosmid 2C4.

FEATURES
source

1. 24225
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2C4"
complement(2. 1477)
/note="Previously sequenced DNA fragment. EMBL:AF113603
Streptomyces coelicolor putative acyl-CoA carboxylase
complex A subunit (accA) gene, complete cds."
2. 119
/note="nominal overlap with Streptomyces coelicolor cosmid
SCA10"
complement(374. 1471)
/note="Pfam match to entry PF00289 CPSase-L_chain,
Cardamoyl-phosphate synthase (CPSase), score 585.30,
E-value 3.8e-172"
complement(602. 625)
/note="PS00867 Cardamoyl-phosphate synthase subdomain
signature 2"
complement(1485. 1489)
1774. 1778
1786. 3438
/gene="SC2C4.02"
/note="SC06272"
1786. 3438
/gene="SC2C4.02"
/note="SC06272"
1786. 3438
/gene="SC2C4.02"
/note="SC2C4.02, possible secreted FAD-binding protein,
len: 550 aa; similar to TR:Q92AR8 (EMBL:U80223)
Streptomyces penicillius ceasius Dnrm, 485 aa; fasta scores:
opt: 1323 z-score: 1315.4 E(): 0: 45.3% identity in 470 aa
overlap. Contains Pfam match to entry PF01565
FAD-binding_4, FAD binding domain. Also contains possible
N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted FAD-binding protein"
/protein_id="CAC22143.1"
/db_xref="GI:12274800"
/db_xref="SPTREMBL:Q9EX55"
/translation="MTEVSRRLKMGAAVSGALALPALGAPATAAPAGPEDLPGR
AAAAAGRRPAPVHVYRSDPRYQGLISNSQNRWNPDIYRVSSAGVYRAQEA
DKGLKLVRSRGGDEDEEVANRDVYIDAMGMSYVDREERAFVAGVRLCTYVR
TLYKRMGVTVLPGCTCPYVAGGHTGGGATLSRGLTYDHLAYAVVYVDAAGRAR
KYVATREDDPNRELMAHRTGAGGAGCGVITTRYLRSPDVGDDPSTLLPRAEVEVL
SDVSWMDDEDEAFTRLRNFTETMARNSAPDSGFRFLSQTLMKRAAGYFEMSTV
VDAAPADRLDLAYLAISGEGTVYHVDYRAMPYLVATEMSGFEVSAVPMKSK
SAVREVAPEQLAVYRQLTRDYPGYPMIAIVGCGKINEVAPGDTATAGDSIA
KMLYGLMSDPADALHQRMTREAVEDVYASTGVPVPGGVNDGCYINTYADADLADPA
LNRSGIPMHLTYFGNTPRLQRYKATMDPRNVPSHRIGTKRP"
1948. 2355
/gene="SC2C4.02"
/note="Pfam match to entry PF01565 FAD-binding_4, FAD
binding domain, score 6.50, E-value 1.1e-06"
complement(3514. 9972)
/gene="SC2C4.03c"
/note="SC06273"
complement(3514. 9972)
/gene="SC2C4.03c"
/note="SC2C4.03c, probable type I polyketide synthase,

len: 2152 aa; N-terminal domain almost identical to
TR:Q918J0 (EMBL:AF202898) Streptomyces coelicolor A3(2)
type I polyketide synthase (fragment), 720 aa; fasta
scores: opt: 4356 z-score: 4494.6 E(): 0: 93.8% identity
in 721 aa overlap and C-terminal domain similar to
SW:HEYM_AANAP (EMBL:L22883) Anabaena sp. polyketide
synthase Hetm, 506 aa; blastp scores: Score=761 (267.9
bits), Expect=2.2e-74, Sum P(2)=2.2e-74; Identifies=184/491 (37%). Positives=278/491 (56%). Also highly
similar, in this same cosmid, to the N-terminal region of
SC2C4.04c and to SC2C4.03c. Contains Pfam matches to
entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase,
PF00698 Acyl-transf, Acyl transferase domain and PF00550
pp-binding, Phosphopantetheine attachment site and matches
to Prosite entries PS00606 Beta-ketoacyl synthases active
site, PS00340 growth factor and cytokines receptors family
signature 2 and PS00012 Phosphopantetheine attachment
site"

/codon_start=1
/transl_table=11
/product="putative type I polyketide synthase"
/protein_id="CAC22144.1"
/db_xref="GI:12274801"
/db_xref="SPTREMBL:Q9EX54"
/translation="MSNEOKMVEYLKWTVDLQKARQRLAELENGNDPEYAIMACR
PFGVVASPDMLRLVMDADGIDTMDPDRMDIDAIVDPERKGRKSYTRKGGFLDGA
AOPDPGPFGISPRALAMDPOQRULETTANETPERQADIDPATLRGSQVAPAVVBS
YLGLEPGEELBGLDMSKLSVASGRATYTFEGEGPASYDTRICSSVALHLAVOSV
RKGESTMALAGGSIVSGNPGSFGFDSKONLIDEGRIKRSRAAADGIVAGGGLLI
EKLSDARRNGRHVLAIVRGSVAVNODGASNGLITPNNGSQVIRHQALAGARLANSDD
VVEAHGTRGLGDIIEAOLALASGRPRADRPRLYGLTNSINGHYVAAAGAGVIGM
IMMRHGILIKTLTVNDEPPLSMENAGAVLLETPARPEPTGAPRAGVAGPISGN
AHVLEERPAEPEERPDYATLAPVLPALASRTAPALADQARLLSHYERPAEDP
LDVAVSLAERSALEHRAVAVVGSREELTLGLRLALAGRTLPAGAVDTQGTATFL
FTGGAGQRTGMDLRYFPFAAHAFDTITAOQDPLHDOPPLHTTTGHLNHTNQ
PALFATVVALYRLLETWGITPDYLAGHSIELRAHISGLITLQDADCTVTAARADQ
NLPHGTMIALOATEOEILPHLTGHEHNLITLAINSPSTSVISGNCTAACHIAATQDA
OGRKTKLTIVSHAFSPHMDGMIDHHTHTAOLITLPHPTPIYVSTLGNLATHNDLT
PIYVTDOLNRYVTRTQAHTHLHAGVITYTTEIGEDATLPELTQITDTPATRLARN
QPEYHTLTLTISRLLHNGIPIPMKNLITFTTPAKRTPLPTAFQHERKWTESGYADA
SIDGLSAGHPEMLGAATVAVAGADVLTSRLSLTHNMLADHVLLDTPALVALVM
AIRAGDELGSVILNDLSVLPLVPGGAVHVOVSPPASGRLRELSISRPDNAV
PALTATNARGRLSFRGPDAPFSLACMPAPGAELEPDEYDLKAVOGYVGYRGLSEY
WRCTDEYVADYRLDPGSRGTGGGFLPALLETALALASCLAGSGSEGAWSTDK
DYRLAHGADRLRYRLAPADGALNOLADRPAPVAVSVSVALRAVGEVSTALAL
NRSLFHYDNAPRPVADPSAGTERVNSGADBPGRKRPDYASLARAVAAESVDVALY
GTEGDTGDRVADHDVTRALALVQDMLAEERLAGTRLVKRGVTTAGDEVDL
GAAVWGLIRSAQSENPGRITLVDDDEEASRLTPIVAVASGEPOSVVRGQVFPRL
GRVPEAPAGATGAMDRRGTVLITGSGSLGGLFARLIVVEHGRNLLTSRGRDAE
GAELLEALFALGAVREITLAODADRAALAVLADIPAEVPLGVLTAGVLDGDLIG
SQPQRLSAVLRKVDATNMLHETLRMDLSAPVLPSSIAVVGPGQSYSAANHFL
DALHEHRSRGLATSLAKGLMSQTSCTGHDBADKRLARSGFRVYESQGPALD
IALTVGRAGVATPMDLTALREOVQVLLSLKARTPVRRTAOAAEYSSLSERLNG
LGDAEGRVAVIGVILGQAAVVLVGHSDATGIDROSFPALCFSTVELNOLSTSG
LRPPTLVLPFHPVVALAEVLEHGEVRLSADGSAAGVGAFAAEVLEADVADAEV
ITPAADKLEILITNGSGLFALFLRDLARTTAYVHCLVGADEAAAMELKKANAEY
RYVMDIDDEPVSIVGLDAPRLDESDDALARTDYVYVNGARVHMLLPETTLKA
SNVTGTEVRLRAKHRTVYHYVSTYGVDPGCRPEVPLRYVDTPPRASPUSGRTK
SKVAAEVOVLEVARGRGLPVSVYRVADISGRVACQCPRTFVWLTLLGLQANSVGR
TEGTFHLIPVDVYSAAITGTSRQGTGTFHFLNOSLALSOCVELLRSLIGLELEY
DMWDTVOVVISGDAVLLPLDLDAFEMMSSTDGFPPIPTAETVAALEGTGIGIPVLTR
ELPEKYVAFVPEBGPAG"
complement(4732. 4935)
/gene="SC2C4.03c"
/note="Pfam match to entry PF00550 pp-binding,
phosphopantetheine attachment site, score 68.50, E-value
2.5e-17"
complement(4810. 4857)
/gene="SC2C4.03c"
/note="PS00012 Phosphopantetheine attachment site"
complement(5248. 5268)
/gene="SC2C4.03c"
/note="PS00340 growth factor and cytokines receptors

misc_feature		family signature 2"	
complement(7342, 1286)		/gene="SC2C4.03c"	
/note="Pfam match to entry PF00698 Acyl-transf, Acyl transferase domain, score 248.50, E-value 9.5e-71"		complement(8596, 9867)	
/gene="SC2C4.03c"			
<hr/>			
Query Match		45.1%	Score 45.6; DB 1; Length 2425;
Best Local Similarity		66.0%	Pred. No. 0.092;
Matches		66; Conservative	0; Mismatches 34; Indels 0; Gaps 0;
<hr/>			
QY	2	GTTCCTCGGGTCTCTGGCTTCTCCAGCCAGCGGCTGATCTGGATGGGTTCAAT	61
Db		21339	GTGGCGCGGTCCTCGGCCACACCGATCC35CTCTGTGACCCCGCGGCGCTTCAG
QY		62	GAGCAGGCGCTCAGCTCTGATGGCGGTG5AGATCCGA
Db		21279	GAGCTGGGCTTGACTGACTGACAGCGCTC3AACTCCGA
<hr/>			
RESULT 15			
LOCUS		AX211705	65140 bp DNA linear PAT 06-SEP-2001
DEFINITION		Sequence 1 from Patent WO0153126.	
ACCESSION		AX211705	
VERSION		AX211705.1	GI:15523937
KEYWORDS		Streptomyces noursei.	
SOURCE		Streptomyces noursei.	
ORGANISM		Bacteria; Filicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.	
REFERENCE		1 (bases 1 to 65140)	
AUTHORS		Zocchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T., Strom,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and Guliksen,O.M.	
TITLE		Gene cluster encoding a nystatin polyketide synthase and its manipulation and utility	
JOURNAL		Patent: WO 0159126-A 1 16-AUG-2001;	
		Ind og Teknisk Naturvitenskapslige Universitet (NO); STIFTELSEN NORD OG TEKNISK FORSKNING VED NORSKE TEKNISKE HOGSKOLE (NO); ALPHABRA AS (NO); Slavnent AS (NO); Zocchev, Sergey Borisovich (NO); Sekurova, Olga Nikolaevna (NO); Fjaervik, Espen (NO); Brautaset, Trygve (NO); Strøm, Arne Reidar (NO); Valla, Svein (NO)	
FEATURES		Location/Qualifiers	
source		1..65140	
		/organism="Streptomyces noursei"	
		/db_xref="taxon:1971"	
		/note="ATCC 11455"	
BASE COUNT		8270 a	25171 c 22273 g 9426 t
<hr/>			
ORIGIN			
Query Match		45.1%	Score 45.6; DB 6; Length 65140;
Best Local Similarity		66.0%	Pred. No. 0.081;
Matches		66; Conservative	0; Mismatches 34; Indels 0; Gaps 0;
<hr/>			
QY	2	GTTCCTCGGGTCTCTGGCTTCTCCAGCCAGCGGCTGATCTGGATGGGTTCAAT	61
Db		49388	GCCGCGGTGTCCTCGGCCACCGACGAGCGATCGGCGGCGGAGACGTTCCAG
QY		62	GAGCAGGCGCTCAGCTCTGATGGCGGTGAGATCCGA
Db		49448	GAGTTGGGCTTGACTGCTGACCGCGGAGGATTCGGA

Search completed: November 5, 2002, 12:24:02
Job time : 603.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400

Perfect score: 101

Sequence: 1 gtcgaaccgtagcagacgct.....cgtcacgtgcgagacgca 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estbma:*
2: em_estbma:*
3: em_estin:*
4: em_estma:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.2	35.8	1084	17	AG077752 Pan trogl
2	34	33.7	1296	17	AG073215 Pan trogl
3	32.4	32.1	905	17	AQ0688230 nxd00776
4	32.2	31.9	694	12	BF422350 lr_ad_142
5	31.8	31.5	300	13	BJ248635 BJ248635
6	31.4	31.1	572	13	BI875065 BI875065

7	31	30.7	484	13	BI336276
8	31	30.7	547	12	BF078123
9	30.8	30.5	413	10	AV630710
10	30.6	30.3	479	14	BQ294770
11	30.6	30.3	523	10	AM566146
12	30.6	30.3	590	17	BH782972
13	30.6	30.3	610	9	A1621801
14	30.6	30.3	879	12	BG755515
15	30.6	30.3	1024	11	AY107409
16	30.4	30.1	253	13	BI478653
17	30.4	30.1	747	14	BQ805689
18	30.2	29.9	493	10	BE363468
19	30.2	29.9	540	13	BI809588
20	30.2	29.9	791	17	AG049188
21	30.2	29.9	1285	17	AG058001
22	30	29.7	270	9	AA215413
23	30	29.7	446	9	AA279201
24	30	29.7	457	9	AA828455
25	30	29.7	501	9	AA827645
26	29.8	29.5	796	17	AZ192927
27	29.8	29.5	823	12	BF984458
28	29.8	29.5	1421	17	AQ012194
29	29.6	29.3	464	14	BQ159769
30	29.6	29.3	542	12	BG274058
31	29.6	29.3	547	12	BG274590
32	29.6	29.3	558	12	BG274373
33	29.6	29.3	558	12	BF292551
34	29.6	29.3	565	14	BQ841584
35	29.6	29.3	587	12	BF291659
36	29.6	29.3	598	12	BF291479
37	29.6	29.3	667	14	BQ841258
38	29.6	29.3	669	14	BQ841464
39	29.6	29.3	680	12	BF291310
40	29.6	29.3	681	14	BQ840803
41	29.6	29.3	683	14	BQ841101
42	29.6	29.3	742	12	BF616609
43	29.6	29.3	743	14	BQ840751
44	29.6	29.3	764	14	BQ840589
45	29.6	29.3	978	12	BG860535

ALIGNMENTS

RESULT 1
AG077752/c 1084 bp DNA linear GSS 03-NOV-2001
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-072F16.R, genomic survey sequence.
ACCESSION AG077752
VERSION AG077752.1 GI:16629554
KEYWORDS
SOURCE
ORGANISM
Pan troglodytes male lymphoblast DNA, clone:lib:PTB Chimpanzee Male
BAC library clone:PTB-072F16.R.
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE
BAC end sequences of library PTB
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1084)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totohi,Y., Watanabe,H. and Sakaki,Y.
TITLE
Direct Submission
JOURNAL
Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT
Clones are derived from the chimpanzee BAC library PTB This BAC end
clone tracking errors.

```

PRIMERS
Sequencing: M13rev
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. 1084
/organism="Pan troglodytes"
/db_xref="taxon:9398"
/clone="PTB-072F15.R"
/sex="male"
/cell_type="Lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      248 a      489 c      307 g      13 t      27 others
ORIGIN

Query Match      35.8%; Score 36.2; DB 17; Length 1084;
Best Local Similarity 62.9%; Pred. No. 4.4;
Matches 56; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY      8  CGATGACAGCGCTGTTGTCGTCGAGCCCGCTGCGAGGAGGTGTCACGCCCTG 67
          || || || || || || || || || || || || || || || || || || ||
DB      674  GCGCGTCTCGCGGGGTTTCTCGCGCGCGCGCGGGGCGGCGGTCTTCGCCGCC 615
          || || || || || || || || || || || || || || || || || || ||
OY      68  GCGTCAGCGCTCATGCTCCACTGCGCGAG 96
          || || || || || || || || || || || || || || || || || || ||
DB      614  GCGGCTTCGCGTTCGCGCGCGCGCGCGCG 586
          || || || || || || || || || || || || || || || || || || ||

RESULT 2
AG073215      1290 bp      DNA      linear      GSS 03-NOV-2001
LOCUS      Pan troglodytes DNA, clone: PTB-064P04.F, genomic survey sequence.
ACCESSION      AG073215.1 GI:16625017
VERSION      GSS.
KEYWORDS      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
SOURCE      BAC Library clone: PTB-064P04.F.
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE      1  Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
AUTHORS      Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      BAC end sequences of Library PTB
JOURNAL      Unpublished
REFERENCE      2  (bases 1 to 1296)
AUTHORS      Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
TITLE      Totoki, Y., Watanabe, H. and Sakaki, Y.
JOURNAL      Direct Submission
COMMENT      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9111, Fax: 81-45-503-9170
Clones are derived from the chimpanzee BAC library PTB this BAC end
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. 1296
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-064P04.F"
/sex="male"
/cell_type="Lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
FEATURES
source

```

```

BASE COUNT      338 a      287 c      531 g      126 t      14 others
ORIGIN

Query Match      33.7%; Score 34; DB 17; Length 1296;
Best Local Similarity 58.6%; Pred. No. 18;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY      1  GTGACACGATGACAGCGCTGTTGTCGTCGAGACCCCGCTGCGAGGGGTGTCAC 60
          || || || || || || || || || || || || || || || || || || ||
DB      1183  GTAGACCGAGACCGCGCGCGCGCTTCTCGCCGACCCCGCAGAGAGAGAGACCG 1242
          || || || || || || || || || || || || || || || || || || ||
OY      61  GCGCGTGGCGTCAGCGTCATGCGTCCACTGCGGAGACG 99
          || || || || || || || || || || || || || || || || || || ||
DB      1243  GCAGGTGCGGTAGGTGACGCGCAGAGAGAGAGAGAGCGC 1281
          || || || || || || || || || || || || || || || || || || ||

RESULT 3
AO688230      905 bp      DNA      linear      GSS 01-JUL-1999
LOCUS      nbxb0077G03r CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION      nbxb0077G03r, DNA sequence.
VERSION      AO688230
KEYWORDS      AO688230.1 GI:5329314
SOURCE      GSS.
ORGANISM      Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE      1  (bases 1 to 905)
AUTHORS      Wing, R.A. and Dean, R.A.
TITLE      A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL      Unpublished (1998)
COMMENT      Contact: Ming RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAACACCTATGACCATG
Class: BAC ends
High quality sequence start: 46
High quality sequence stop: 195.
Location/Qualifiers
1. 905
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbxb0077G03r"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="leaf"
/lab_host="E. coli DH10b"
/notes="Vector: pBeloBAC11, Site 1: HindIII, Site 2:
HindIII. Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."
FEATURES
source

```

BASE COUNT	154 a	302 c	267 g	177 t	5 others
ORIGIN					
Query Match					
Best Local Similarity	32.1%;	Score 32.4;	DB 17;	Length 905;	
Matches 51;	Conservative 0;	Mismatches 31;	Indels 0;	Gaps 0;	
OY	12	GACAGCGTGGTTTGTCGTGCGAGACCCCGCTGCGAGAGGGTGGTGCACGCCCGCTGGCGGT	71		
Db	70	GGCGGCGCTGTAGGAGGCGGCGCTGCTCCCGAGCGCGGAGAGGTCCTCTACCCGCGGT	129		
OY	72	CAGCGTCATGGCGTCACATGGCG	93		
Db	130	CCCCGTGGCGGCTGAACTTGGC	151		
RESULT 4					
LOCUS 232350/c					
DEFINITION	BF423350	694 bp	mRNA	linear	EST 28-NOV-2000
ACCESSION	BF423350				
VERSION	BF423350.1	GI:11410339			
KEYWORDS	EST.				
SOURCE	humus earthworm.				
ORGANISM	Lumbricus rubellus				
REFERENCE	Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxidea;				
AUTHORS	Lumbricina; Lumbricidae; Lumbricus.				
TITLE	1 (bases 1 to 694)				
JOURNAL	Sturzenbaum S., Parkinson J., Blaxter M., Morgan J., Kille P.,				
COMMENT	Schallinger W., and Georgiev O.				
	Expressed Sequence Tags from the humus earthworm L. rubellus				
	Unpublished (2000)				
	Contact: Sturzenbaum SR				
	School of Biosciences				
	Cardiff University				
	PO Box 911, Biomedical Building, Museum Avenue, Cardiff, CF10 3US,				
	UK				
	Tel: +44 2920 874119				
	Fax: +44 2920 874094				
	Email: SturzenbaumSR@cardiff.ac.uk				
	Seq primer: T3 sequencing primer				
	High quality sequence stop: 487.				
FEATURES					
source	Location/Qualifiers				
	1..694				
	/organism="Lumbricus rubellus"				
	/db_xref="taxon:35632"				
	/clone="Lr_ad_142"				
	/clone_lib="Earthworm Lambda Zap Express Library"				
	/rclone_type="whole worm"				
	/dev_stage="adult"				
	/note="Vector: pBK-CMV. The library was prepared using				
	protocol given by supplier (Stratagene)."				
BASE COUNT	159 a	172 c	211 g	152 t	
ORIGIN					
Query Match					
Best Local Similarity	31.9%;	Score 32.2;	DB 12;	Length 694;	
Matches 58;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0;	
OY	1	GTCGAACGATGACAGCGCTGATTTCGTGCGTGCAGCCCGCTGCGAGAGGGGTGTGCAC	60		
Db	564	GTCACACCTTTAAGAAATCCGACATCCGACGATGCGCTCGCGCATGACTGATGATGTAC	505		
OY	61	GCCGCTGGCGTCAGCGTCATGCGCTCACTGGCGGAGACGA	101		
Db	504	AGCGGAGCCCTCAGACGATGTAATTCGTGGGAGACGAGAA	464		
RESULT 5					
LOCUS	BU248635	300 bp	mRNA	linear	EST 05-APR-2002

DEFINITION	Bj248635 Y. Ogihara unpublished cDNA library, Wh_f Triticum aestivum CDNA clone whf7e20 5', mRNA sequence.
.ACCESSION	Bj248635
VERSION	Bj248635.1 GI:20059371
KEYWORDS	EST.
SOURCE	bread wheat.
ORGANISM	Triticum aestivum
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.
AUTHORS	Ogihara,Y and Murai,K
JOURNAL	Expressed genes in Triticum aestivum unpublished (2002)
COMMENT	Contact: Tadasu Shin-1 Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshin@genes.nig.ac.jp. location/Qualifiers
FEATURES	1..300
source	/organism="Triticum aestivum" /cultivar="Chinese Spring" /db_xref="taxon:4565" /clone="whf7e20" /clone_lib="Y. Ogihara unpublished cDNA library, Wh_f" /tissue_type="spike at flowering date" /dev_stage="freezes' scale 10.5.1" /note=Vector: Lambda Xht-LAP XR, excised phagemid: Site_1: EcoRI; Site_2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J Dvorak Lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give plasmid script phagmids in the TJ Clote lab at the University of California, Riverside (Akhunov, Chin , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang), Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors).
BASE COUNT	62 a 89 c 86 g 60 t 3 others
ORIGIN	
Query Match	31.5%; Score 31.8; DB 13; Length 300;
Best local similarity	56.4% Pred.No.57; Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Dy	1 GTGCAACCGATGACAGCCCTGTTCGTGGTGACGCCCGCTGCAGAGGGGTGTGCAC 60 Db 6 GGCAAGCGCGCNGNNGSTGGCTTCGTGGCGCGCGCGCTGCCGGGGAGGTGACA 65 Oy 61 GCCGCTGGCGTCAGCGTCATGCGCTCACACTGCGGAGACGA 101 Db 66 GCCGCGCGCGCGCATCTCTCCCCCGCGCGCGGCTGCA 106
RESULT 6	
LOCUS	B1875065 572 bp mRNA linear EST 11-OCT-2001
DEFINITION	96312E12.y1 C. reinhardtii CC-1690, stress condition I, normalized
ACCESSION	B1875065
VERSION	B1875065.1 GI:16073069
KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.
ORGANISM	Chlamydomonas reinhardtii Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE	1 (bases 1 to 572) Grossman,A., Chang,C.-W., Davies,J.J., Harris,E., Hauser,C., Lefebvre

TITLE
P.J. McDermott, J.P. Shrager, J. Silflow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 963
JOURNAL
Unpublished (2001)
COMMENT
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source
1. 572
Location/Qualifiers

/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3455"
/clone_lib="C. reinhardtii CC-1690, stress condition 1,
normalized, lambda Zap II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExSist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT
ORIGIN
115 a 159 c 189 g 109 t

Query Match
Best Local Similarity 31.1%; Score 31.4; DB 13; Length 572;
Matches 47; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 24 TTGCTGGTGCAGCCCCCGCTGGAGGGGTGTCACCCCGCTGGCTCATGCC 83
DB 124 TTCTCTCGCGGGGCTGTTCAGCAGCTGTGTGTCAGCGGCTTATTCG 183
QY 84 TCCACTGCGCGAG 96
DB 184 TCAGATGCTGAG 196

RESULT 7
B1336276 484 bp mRNA linear EST 01-SEP-2001
LOCUS B1336276
DEFINITION AR094C04SPC04S Porcine spleen cDNA library Sus scrofa cDNA, mRNA
sequence.
ACCESSION B1336276
VERSION B1336276.1 GI:15417572
KEYWORDS EST.
SOURCE P19.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 484)
Rink, A., Santachi, E.M. and Beattie, C.W.
Amplified, Normalized cDNA Libraries from a Porcine Model of
Orthopedic Implant Associated Staphylococcus aureus Infection
Unpublished (2001)
CONTACT: Rink A
Department of Animal Biotechnology
College of Agriculture, Biotechnology and Natural Resources,
University of Nevada, Reno
MS 202, FA 103, 1664 N Virginia St, Reno, NV 89557-0236, USA
Tel: 775 784 1705
Fax: 775 784 1375
Email: arink@cabnr.unr.edu

Tissues and cells are derived from a porcine model for
implant-associated infection using 1000 cfu of Staphylococcus
aureus in a tibial transection. Reduced and internally fixed with a
dynamic compression plate. NOTE: The sequences contain a 'cDNA
adapter' between the EcoRI site and the start of the EST. The
adapter sequence is 'AATTGCGACGAG'.

FEATURES
source
1. 484
Location/Qualifiers

/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone_lib="Porcine Spleen cDNA library"
/tissue_type="Spleen"
/cell_type="mixed"
/dev_stage="control, 5 month old castrated male"
/lab_host="SOLR"
/note="Vector: pBSK; Site 1: Eco RI; Site 2: XhoI; Tissues
and cells are derived from a porcine model for
implant-associated infection using 1000 cfu of
Staphylococcus aureus in a tibial transection, reduced and
internally fixed with a dynamic compression plate. NOTE:
The sequences contain a 'cDNA adapter' between the EcoRI
site and the start of the EST. The adapter sequence is
'AATTGCGACGAG'."

BASE COUNT
ORIGIN
90 a 138 c 158 g 82 t 16 others

Query Match
Best Local Similarity 30.7%; Score 31; DB 13; Length 484;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 2 TCGACCGATGACAGCGCTGTTCCTGCTGAGACCCCGCTCGAGGGGTGTCAGC 61
DB 137 TGGGCCCGACAGACAGAGCTGTGTGCGCGGGCGCCCTGCTGTGTGCGGCGCG 196
QY 62 CCGCTGCGCTGACGCTGATGCTGCAC 88
DB 197 CGCGCGCGCTGTGTGCTGCTGCTGCC 223

RESULT 8
BF078123 547 bp mRNA linear EST 18-OCT-2000
LOCUS BF078123
DEFINITION 228383 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF078123
VERSION BF078123.1 GI:10871953
KEYWORDS EST.
SOURCE P19.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 547)
Fahrenkrug, S.C., Freiling, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.W., Bennett, G.A., Laegreid, W.W.
and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
CONTACT: Smith RPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithr@mail.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTACAGAGC
Plate: 49 row: 3 column: 24
Seq primer: AATTGAGTACACTATAG.
Location/Qualifiers

LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL MEDLINE	COMMENT	FEATURES	BASE COUNT	ORIGIN
AV630710/c	AV630710 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone lcl083a07_r 5', mRNA sequence.	AV630710	AV630710.1	GI:10793344	Chlamydomonas reinhardtii.	1 (bases 1 to 413)	Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohgama, K., Nakamura, Y. and Tabata, S.	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii	DNA Res. 7 (5), 305-307 (2000)	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers	1. .413 /organism="Chlamydomonas reinhardtii" /strain="C9" /db_xref="taxon:3055" /clone="lcl083a07_r" /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"	75 a 135 c 152 g 51 t	1. .547 /organism="Sus scrofa" /db_xref="taxon:9823" /clone_lib="MARC 2Pig" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPORT6; site 1: NotI; site 2: SalI; library made from pooled tissue from testis, cVary, endometrium, hypothalamus, pituitary, and placenta."
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;								
Db	Query Match	30.7%	Score 31;	DB 12;	Length 547;								
Db	Best Local Similarity	59.8%	Pred. NO. 1e+02;										
Db	Matches 52;	Conservative 0;	Mismatches 35;	Indels 0;	G								

[illegible]

DEFINITION 660062P05.y1 660 - Mixed stages of anther and pollen Zea mays cDNA, mRNA sequence.

ACCESSION AM566146

VERSION AM566146.1 GI:7227505

KEYWORDS EST

SOURCE Zea mays.

ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

AUTHORS 1 (bases 1 to 523)

TITLE Walbot V.

JOURNAL Maize ESTs from various cDNA libraries sequenced at Stanford University

COMMENT Unpublished (1999)

CONTACT Walbot V

DEPARTMENT Department of Biological Sciences

STANFORD UNIVERSITY

855 California Ave, Palo Alto, CA 94304, USA

TEL: 650 723 2227

FAX: 650 725 8221

EMAIL: walbot@stanford.edu

PLATE: 660062 row: F column: 05.

FEATURES

source 1..523

/organism="Zea mays"

/cultivar="Ohio43"

/db_xref="taxon:4577"

/clone_lib="660 - Mixed stages of anther and pollen"

/tissue_type="Whole premeiotic anthers to pollen shed"

/dev_stage="premeiotic anthers to pollen shed"

/lab_host="XILOR"

/note="Organ: anthers; Vector: Lambda zap; Site_1: EcoRI; Site_2: XhoI; Anther and pollen cDNA library. Directionally sequenced with 5' end at the EcoRI site. Created by Amie Franklin."

BASE COUNT 105 a 156 c 131 g 130 t 1 others

ORIGIN

Query Match 30.3%; Score 30.6; DB 10; Length 523;

Best Local Similarity 58.1%; Pred. No. 1.3e+02;

Matches 54; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 7 CCGATGACAGCGCTGTTCTGTCGTCAGCCCCCGCTGCGAGGGGTGTCAGCGCGCT 66

DB 111 CCGATGACAGCGCGCTGTCGTCAGCGCGAGAGCGCTGTCGAGCGAGCGCGCTCC 52

QY 67 GCGGTGAGGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99

DB 51 GCGGCCAGCGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19

RESULT 12

BH782972 590 bp DNA linear GSS 28-MAR-2002

LOCUS BH782972/c

DEFINITION fmb011f017h01k0 fmb filtered library Zea mays genomic clone

ACCESSION fmb011f017h01 5, DNA sequence.

VERSION BH782972

KEYWORDS BH782972.1 GI:19786407

SOURCE GSS.

ORGANISM Zea mays.

Ze mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 590)

AUTHORS Budiman, M.A., Freese, R.G., Betell, J.A., Nunberg, A.N. and Lakey, N.D.

TITLE Genethesher methylation filtered genomic sequences from maize

JOURNAL Unpublished (2002)

CONTACT: Bedell JA

ORION GENOMICS, LLC

4041 Forest Park Ave, St. Louis, MO 63108, USA

TEL: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Plate: fmb011f017 row: h column: 01

Seq primer: SK reverse

Class: Shotgun

High quality sequence stop: 590.

FEATURES

source 1..590

/organism="Zea mays"

/cultivar="M017"

/db_xref="taxon:4577"

/clone_lib="fmb011f017h01"

/note="Organ: leaf; Vector: pBSK(-); Site_1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBSK(-) vector and electroporated into E. coli cells."

BASE COUNT 98 a 210 c 176 g 106 t

ORIGIN

Query Match 30.3%; Score 30.6; DB 17; Length 590;

Best Local Similarity 56.4%; Pred. No. 1.3e+02;

Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1 GTGCAACCATGACACCGCTGTTCTGTCGTCGACGCCCGCTGCGAGGGTGTGTCAC 60

DB 242 GTGCTACTGTCGCGACGCGCGGTTCCGCGGCTCCAGACCTCAGGAGAGAC 183

QY 61 GCCGCTGCGTCAGCGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101

DB 182 GCGCCGCGGCGCTCAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 142

RESULT 13

A1621801 610 bp mRNA linear EST 22-APR-1999

LOCUS A1621801

DEFINITION 486026B10.x1 486 - leaf primordia cDNA library from Hake lab Zea mays cDNA, mRNA sequence.

ACCESSION A1621801

VERSION A1621801.1 GI:4646726

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 610)

AUTHORS Walbot V.

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University

COMMENT Unpublished (1999)

CONTACT Walbot V

DEPARTMENT Department of Biological Sciences

STANFORD UNIVERSITY

855 California Ave, Palo Alto, CA 94304, USA

TEL: 650 723 2227

FAX: 650 725 8221

EMAIL: walbot@stanford.edu

PLATE: 486026 row: B column: 10.

FEATURES

source 1..610

/organism="Zea mays"

/cultivar="B73"

/db_xref="taxon:4577"

/clone_lib="486 - leaf primordia cDNA library from Hake lab"

/tissue_type="leaf primordia"

/dev_stage="P7-p11 leaf"

/lab_host="E.coli XL1-Blue MFR"

/note="Organ: Shoot; Vector: Lambda zap; Hake lab cDNA library."

BASE COUNT 82 a 226 c 189 g 107 t 6 others

Query Match 29.3% Score 29.6; DB 10; Length 2787;
Best Local Similarity 57.6%; Pred. No. 2.9;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 2 TCGAACCATGACAGCGCTGTTTCGTGCGTCGAGCCCCCGCTGCGAGGGGTGTGCACG 61
DB 275 TCGAACCCCTTGCACCGCATGAGTTAAAGCGATGAGACTGCCCTGCGGTTCTGCGG 334
OY 62 CCGTCGGCTCAGCGTCATGCTCCACTG 93
DB 335 TTAGAGCGGACGACGCTATGCGTACTCTG 366

RESULT 2
US-09-923-876-949/c
; Sequence 949, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 949
; LENGTH: 258
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700157856H1
; NAME/KEY: unsure
; LOCATION: 17, 38, 52, 75, 85, 92, 101, 115, 175, 196, 239
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-949

Query Match 28.3% Score 38.6; DB 10; Length 258;
Best Local Similarity 54.2%; Pred. No. 5.1;
Matches 52; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 5 AACCATGACAGCGCTGTTTCGTGCGTCGAGCCCCCGCTGCGAGGGGTGTGCACGCG 64
DB 166 ACCCCCGCCCTCGGTGGCGCGCTCGAGCAGCCGCGCTGCTGTGGCMGCGCGAG 107
OY 65 CTGCGTCAGCGTCATGCTCCACTGCGG/GAGCG 100
DB 106 CGAGNCATGTCATGATGACNGCTGGCGG 71

RESULT 3
US-09-954-456-675/c
; Sequence 675, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923

; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 675
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: n-a,t,g or c
US-09-954-456-675

Query Match 27.7% Score 28; DB 10; Length 336;
Best Local Similarity 64.5%; Pred. No. 7.5;
Matches 40; Conservative 1; Mismatches 21; Indels 0; Gaps 0;

OY 34 GAGCCCCGCTCGAGGGGTGTGCACCCCGCTGCGCTCAGCGTCATCTCCACTGCGG 93
DB 301 GAGCCAGGCGCCCGAGTGAAGTCTCGAGTGAAGAGGAGGAGTGGCGCGGCC 242
OY 94 GA 95
DB 241 GA 240

RESULT 4
US-09-880-107-1676/c
; Sequence 1676, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Scherf, Joseph G.
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1676
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D51133
; NAME/KEY: unsure
; LOCATION: (1)..
; OTHER INFORMATION: n - a or c or g or t
US-09-880-107-1676

Query Match 27.7% Score 28; DB 10; Length 336;
Best Local Similarity 64.5%; Pred. No. 7.5;
Matches 40; Conservative 1; Mismatches 21; Indels 0; Gaps 0;


```

QY      61  GCCGCTGGCGTCAGCGTCATGCGTCCAC1TGGCGGA  95
          ||||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      175  GCCGACGCGTCTCCGCGCTGCGCGCGC1TGGCGGA  209

```

RESULT 8
US-09-923-876-2583/c

Query Match	27.1%;	Score 27.4;	DB 10;	Length 254;
Best Local Similarity	55.3%;	Pred. No. 11;		
Matches 52;	Conservative 0;	Mismatches 42;	Indels 0;	Gaps 0

```

RESULT 9
US-09-822-849A-526/C
: Sequence 526, Application US/09822849A
: Patent No. US20020045170A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Wong, Gordon G.
: APPLICANT: Clark, Hilary
: APPLICANT: Fechtel, Kim
: APPLICANT: Agostino, Michael J.
: APPLICANT: Howes, Steven H.
: APPLICANT: Resnick, Richard J.
: APPLICANT: Gulikota, Kamalakart
: APPLICANT: Graham, James R.
: APPLICANT: Genetics Institute, Inc.
:
: TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
:
: FILE REFERENCE: GIN 6403
:
: CURRENT APPLICATION NUMBER: US/09/822,849A
: CURRENT FILING DATE: 2001-09-04
: PRIOR APPLICATION NUMBER: 60/195,582
: PRIOR FILING DATE: 2000-04-06
:
: NUMBER OF SEQ ID NOS: 598
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 526
:
: LENGTH: 2084
:
: TYPE: DNA

```

ORGANISM: Homo sapiens
US-09-822-849A-526

Query Match	27.1%	Score 27.4;	DB 10;	Length 2084;
Best Local Similarity	57.6%	Pred. No. 12;		
Matches 49;	Conservative 0;	Mismatches 36;	Indels 0;	Gaps 0;

```

RESULT 10
US-09-815-242-4163
Sequence 4163, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-32
PRIOR APPLICATION NUMBER: 60/265,308
PRIOR FILING DATE: 2001-02-16
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 4163
LENGTH: 2742
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4163

```

	Query Match	27.1%	Score 27.4	DB 10	Length 2742
	Best Local Similarity	55.9%	Pred. No. 12		
	Matches 52	Conservative 0	Mismatches 41	Indels 0	Gaps
QY	1	GTCAACCGAGACAGACGGCTGTTGTCGCGGTGACAGCCCGCGTGGACGAGGGGTGTCAC	60		
DB	265	GTCAAGCCCGTCACATGCCAGCGTCCGCGCCCTCGCGCTTGTCTGTGTGGAGGGG	32		
QY	61	GCCGCTGCGCTACCGCTCATCGCTTCACACTGGCG	93		
DB	325	GTTCAGAGCGCCGACGACGTGATCGGACACCTGGCG	357		

RESULT 11
US-09-923-876-377/c
; Sequence 377, Application US/09923876
; Patent No. US20020013958A1

```
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Itc)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923, 876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 377
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700156836H1
US-09-923-876-377
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 216;
Best Local Similarity 55.2%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```

```
OY 5 AACGATGACAGCGCTGTTGTCGTCGAGCCCGCTCGAGGGGTGTCACGCCG 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 216 AGCCGCCGCCCTCGGTGCGCCCTCTGAGACCGCGCGCTTGCTGTCGCGCG 157
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 65 CTGGCGTCAGCGTCATGCGTCACGTCGCGGAGACGG 100
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 156 CGAGACCGGTGCTTGTATGCTGCTGCGCGCGCG 121
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 12
US-09-923-876-336/c
; Sequence 336, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Itc)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923, 876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 336
; LENGTH: 257
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700156768H1
; NAME/KEY: unsure
; LOCATION: 215, 240, 244, 246
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-336
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 257;
Best Local Similarity 56.2%; Pred. No. 12;
Matches 50; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
OY 12 GACAGCGCTGTTGTCGTCGAGCCCGCTCGAGGGGTGTCACGCCGCTGCGCT 71
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 236 GCCCTGCGTGGCGCTCGAGANCCGCCGCGCTTGCTGCTGCTGCGCGAGAGAC 177
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
OY 72 CAGCGTCACGTCACACTGCGGAGACGG 100
    || ||| || | ||||| ||| |||
DB 176 CATGCTTGTATGCTGCTGCGCGCGCG 148
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 13
US-09-923-876-3689/c
; Sequence 3689, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Itc)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923, 876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 3689
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700405434H1
US-09-923-876-3689
```

```
Query Match          26.9%; Score 27.2; DB 10; Length 265;
Best Local Similarity 55.2%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```

```
OY 5 AACGATGACAGCGCTGTTGTCGTCGAGCCCGCTCGAGGGGTGTCACGCCG 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 216 AGCCGCCGCCCTCGGTGCGCCCTCTGAGACCGCGCGCTTGCTGTCGCGCG 157
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 65 CTGGCGTCAGCGTCATGCGTCACGTCGCGGAGACGG 100
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 156 CGAGACCGGTGCTTGTATGCTGCTGCGCGCGCG 121
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 14
US-09-923-876-3703/c
; Sequence 3703, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Itc)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923, 876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 3703
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700405458H1
US-09-923-876-3703
```

Query Match 26.9%; Score 27.2; DB 10; Length 274;
Best local Similarity 55.2%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 5 AACCATGACAGCGCGTGTTCGTCGATCGACGCCCGCTGCCAGGGGTGTCAGCGCCG 64
DB 217 AGCCGCGCGCGCTGCGCGCGCGCTCGACAGCGCGCGCGCGCTGCGCGCGCG 158

QY 65 CTGCGCTCAGCGCTCATGCGTCACATGCGC3GAGACGG 100
DB 157 CGAGGACCGCTGTCTGTGATGCTGCTGCGC3GCGGCGG 122

RESULT 15
US-09-864-761-27835
; Sequence 27835, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00561
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00570
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,637
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27835
; LENGTH: 361
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003042.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 0.68
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL - 0.74

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL - 0.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 0.83
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL - 0.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL - 0.71
; OTHER INFORMATION: EST_HUMAN HIT: B273103.1, EVALUE 1.00e-112
; OTHER INFORMATION: SWISSPROT HIT: P39290, EVALUE 1.00e-06
US-09-864-761-27835

Query Match 26.9%; Score 27.2; DB 10; Length 361;
Best local Similarity 61.1%; Pred. No. 13;
Matches 44; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 30 GGTGAGACCCCGCTGCGAGGGGTGTCGACGCCCGCTGCGAGCGTCATGCGTCACT 89
DB 87 GAGAGAGCCACACTGCTGGGGTCTGCTGCGCTGCGCTGCCCGCTCTCTCAAG 146

QY 90 GCGGAGACGGA 101
DB 147 GCGGAGGCGCA 158

Search completed: November 5, 2002, 23:00:16
Job time : 83.2537 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400

Perfect score: 101

Sequence: 1 gtcgaaccgatgacagcgct.....cgtccactgctgcgagacgga 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/lna/backfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	68750	3	US-09-335-409-1
2	101	100.0	68750	4	US-09-568-102-1
3	101	100.0	68750	4	US-09-567-969-1
4	101	100.0	68750	4	US-09-568-480-1
5	101	100.0	68750	4	US-09-568-486-1
6	101	100.0	68750	4	US-09-568-472-1
7	101	100.0	68750	4	US-09-567-899-1
8	101	100.0	68750	4	US-09-443-501A-2
9	29.6	29.3	33529	4	US-09-144-085-3
10	29	28.7	44377	2	US-08-804-227C-7
11	29	28.7	44377	2	US-08-804-198-1
12	28.6	28.3	80161	3	US-09-036-987A-1
13	28.6	28.3	80161	3	US-09-370-700-1
14	28.4	28.1	13987	2	US-08-804-227C-13
15	28	27.7	1485	4	US-09-372-422A-39
16	28	27.7	28958	1	US-08-258-261B-6
17	28	27.7	28958	1	US-08-456-837-6
18	28	27.7	28958	1	US-08-457-342-6
19	28	27.7	28958	1	US-08-457-646A-6
20	28	27.7	28958	1	US-08-458-076A-6
21	28	27.7	28958	1	US-08-764-233A-4
22	28	27.7	28958	1	US-08-457-335A-6
23	28	27.7	28958	1	US-08-729-214-6
24	28	27.7	28958	1	US-09-028-934-6
25	28	27.7	49377	1	US-08-764-233A-1
26	28	27.7	4411529	4	US-09-103-840A-1
27	26.8	26.5	1332	1	US-08-660-765A-3

28	26.4	26.1	50937	4	US-09-428-517-1	Sequence 1, Appl1
29	26.2	25.9	667	3	US-09-010-809-13	Sequence 13, Appl1
30	26.2	25.9	2770	3	US-08-586-165-6	Sequence 6, Appl1
31	26.2	25.9	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
32	26	25.7	585	4	US-09-516-914-10	Sequence 10, Appl1
33	26	25.7	14272	4	US-09-516-914-23	Sequence 23, Appl1
34	26	25.7	43280	2	US-08-804-227C-1	Sequence 1, Appl1
35	25.8	25.5	1434	4	US-09-434-288-3	Sequence 3, Appl1
36	25.8	25.5	20235	1	US-07-642-734C-3	Sequence 3, Appl1
37	25.8	25.5	20235	3	US-08-439-009A-3	Sequence 3, Appl1
38	25.6	25.3	8438	1	US-07-945-283-1	Sequence 1, Appl1
39	25.4	25.1	532	4	US-09-449-218D-15	Sequence 15, Appl1
40	25.4	25.1	10095	3	US-08-822-586-45	Sequence 45, Appl1
41	25.4	25.1	49272	1	US-08-614-770A-1	Sequence 1, Appl1
42	25.2	25.0	1851	1	US-08-455-543A-10	Sequence 10, Appl1
43	25.2	25.0	1851	2	US-08-193-078B-10	Sequence 10, Appl1
44	25.2	25.0	1851	2	US-08-223-305C-10	Sequence 10, Appl1
45	25.2	25.0	1851	2	US-08-149-097D-10	Sequence 10, Appl1

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligdon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1

Query Match      100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GTCGAACCGATGACAGCGCTGGTTCTGTCGTCGAGCCCGCTGCGAGAGGAGTGTCAC 60
          |||||||
Db      29912  GTCGAACCGATGACAGCGCTGGTTCTGTCGTCGAGCCCGCTGCGAGAGGAGTGTCAC 29971

QY      61  GCCCGTGGCGTACGCTGATGCGTCCACTGCGGAGAGCGGA 101
          |||||||
Db      29972  GCCCGTGGCGTACGCTGATGCGTCCACTGCGGAGAGCGGA 30012

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligdon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
```

```
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match
  100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3,1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
DB 29912 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 29971

OY 61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
DB 29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012

RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match
  100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3,1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
DB 29912 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 29971

OY 61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
DB 29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012

RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6353458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
```

```
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match
  100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3,1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
DB 29912 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 29971

OY 61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
DB 29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012

RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6353459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match
  100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3,1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
DB 29912 GTCGAACCGATGACAGCCGCTGTTTCGTGCTGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 29971

OY 61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
DB 29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012

RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
```

```

: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: PRIOR FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match
Best Local Similarity 100.0%; Score 101; DB 4; Length 68750;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAACCGATGACAGCCCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCAC 60
Db 29912 GTCGAACCGATGACAGCCCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCAC 29971

QY 61 GCCGCTGGCGTCAGCGTCATGCTCAGTGGCGGAGACGGA 101
Db 29972 GCCGCTGGCGTCAGCGTCATGCTCAGTGGCGGAGACGGA 30012

RESULT 7
US-09-567-899-1
: Sequence 1, Application US/09567899
: Patent No. 6383787
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,899
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match
Best Local Similarity 100.0%; Score 101; DB 4; Length 68750;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAACCGATGACAGCCCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCAC 60
Db 29912 GTCGAACCGATGACAGCCCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCAC 29971

QY 61 GCCGCTGGCGTCAGCGTCATGCTCAGTGGCGGAGACGGA 101
Db 29972 GCCGCTGGCGTCAGCGTCATGCTCAGTGGCGGAGACGGA 30012

RESULT 8
US-09-443-501A-2
: Sequence 2, Application US/09443501A
: Patent No. 6303342
: GENERAL INFORMATION:
```

```

: APPLICANT: Kosan Biosciences, Inc.
: APPLICANT: Julien, Bryan
: APPLICANT: Katz, Leonard
: APPLICANT: Khosla, Chaitan
: APPLICANT: Tang, Li
: APPLICANT: Ziermann, Rainer
: TITLE OF INVENTION: Recombinant Methods and Materials for Producing
: FILE REFERENCE: 30062-20031.00
: CURRENT APPLICATION NUMBER: US/09/443,501A
: CURRENT FILING DATE: 1999-11-19
: PRIOR APPLICATION NUMBER: US 60/130,560
: PRIOR FILING DATE: 1999-04-22
: PRIOR APPLICATION NUMBER: US 60/122,620
: PRIOR FILING DATE: 1999-03-03
: PRIOR APPLICATION NUMBER: US 60/119,386
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: US 60/109,401
: PRIOR FILING DATE: 1998-11-20
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 71989
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match
Best Local Similarity 100.0%; Score 101; DB 4; Length 71989;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAACCGATGACAGCCCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCAC 60
Db 24300 GTCGAACCGATGACAGCCCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCAC 24359

QY 61 GCCGCTGGCGTCAGCGTCATGCTCAGTGGCGGAGACGGA 101
Db 24360 GCCGCTGGCGTCAGCGTCATGCTCAGTGGCGGAGACGGA 24400

RESULT 9
US-09-144-085-3
: Sequence 3, Application US/09144085
: Patent No. 6280999
: GENERAL INFORMATION:
: APPLICANT: Gustafsson, Claes
: APPLICANT: Betlach, Mary C.
: APPLICANT: Ashley, Gary
: APPLICANT: Julien, Bryan
: APPLICANT: Ziermann, Rainer
: TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
: FILE REFERENCE: 30062-20020.20
: CURRENT APPLICATION NUMBER: US/09/144,085
: CURRENT FILING DATE: 1998-08-31
: EARLIER APPLICATION NUMBER: 09/010,809
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 3
: LENGTH: 33529
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-144-085-3

Query Match
Best Local Similarity 29.3%; Score 29.6; DB 4; Length 33529;
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 2 TCGAACCGATGACAGCGCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCACG 61
Db 29972 TCGAACCGATGACAGCGCTGTTTCGTCGTCGAGCCCGCTCGAGGGGTGTCACG 30012
```

Db 24388 TCGGGGCGCTGCTCGCGAGCTGGAGCAGATGAAAGCCCGCTGAGGGCGGTGGCGCACC 24447

Qy 62 CCGCTGGCGTCAGCGTCATCGCTCCACTGGCGGAGACGGA 101

Db 24448 TCGGGGCGATGAGCCCGCGCTGTTCCGCTGGAGAGCTCGA 24487

RESULT 10

US-08-804-227C-7

; Sequence 7, Application US/08804227C;

; Patent No. 5876991

; GENERAL INFORMATION:

; APPLICANT: Dehoff, Bradley S.

; APPLICANT: Kunstoss, Stuart A.

; APPLICANT: Rostock, Paul R., Jr.

; APPLICANT: Sutton, Kimberly L.

; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: THOMAS G. PLANT 1501

; STREET: LILLY CORPORATE CENTER

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: USA

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: ASCII(DOS) Text only

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,27C

; FILING DATE: February 21, 1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Plant, Thomas, G.

; REGISTRATION NUMBER: 35,784

; REFERENCE/DOCKET NUMBER: X-8231

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-2459

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 44377 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 350..14002

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 14046..20036

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 20110..31284

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 31329..36071

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 36155..41830

; US-08-804-227C-7

Query Match 28.7%; Score 29; DB 2; Length 44377;

Best Local Similarity 55.4%; Pred. No. 9.9;

Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1 GTGGAACCGATGACAGCGCTGGTTTGTGCTGAGACCCCGCTGGAGGGGTGTGCAC 60

Db 35820 GGGGGCGCGCTGCTGAGAGCGCTGCTCCGCGGAGAAACCGCGCGGGGGCGCGGAG 35879

Qy 61 GCCGCTGGCGCTCAGCGTCATGCTCCTGCTGAGAGAGCGA 101

Db 35880 CCGCGGGGGGAGAGCCCGGACCCACCGCCCGGAGAGCGA 35920

RESULT 11

US-08-804-198-1

; Sequence 1, Application US/08804198

; Patent No. 5945320

; GENERAL INFORMATION:

; APPLICANT: Burgett, Stanley G.

; APPLICANT: Kunstoss, Stuart A.

; APPLICANT: Rao, Nagaraja R.

; APPLICANT: Richardson, Mark A.

; APPLICANT: Rostock, Paul R., Jr.

; TITLE OF INVENTION: PLANTOLIDE SYNTHASE GENE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PAUL R. CANTRELL 1138

; STREET: LILLY CORPORATE CENTER

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: USA

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Macintosh

; OPERATING SYSTEM: Macintosh 7.0

; SOFTWARE: Microsoft Word 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,198

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Cantrell, Paul R.

; REGISTRATION NUMBER: 36,470

; REFERENCE/DOCKET NUMBER: P9113

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-3885

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 44377 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 350..14002

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 14046..20036

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 20110..31284

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 31329..36071

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 36155..41830

; US-08-804-198-1

Query Match 28.7%; Score 29; DB 2; Length 44377;

Best Local Similarity 55.4%; Pred. No. 9.9;

Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1 GTGGAACCGATGACAGCGCTGGTTTGTGCTGAGACCCCGCTGGAGGGGTGTGCAC 60

Db 35820 GGGGGCGCGCTGCTGAGAGCGCTGCTCCGCGGAGAAACCGCGCGGGGGCGCGGAG 35879

Qy 61 GCCGCTGGCGCTCAGCGTCATGCTCCTGCTGAGAGAGCGA 101

Db 35880 CCGCGGGGGGAGAGCCCGGACCCACCGCCCGGAGAGCGA 35920

```
RESULT 12
US-09-036-987A-1
: Sequence 1, Application US/09036987A
: Patent No. 6143526
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard H.
: APPLICANT: Broughton, Mary C.
: APPLICANT: Crawford, Kathryn P.
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Merlo, Donald J.
: APPLICANT: Treadway, Patti J.
: APPLICANT: Turner, Jan R.
: APPLICANT: Waldron, Clive
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
: TITLE OF INVENTION: Production
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dow Agrosciences LLC Patent Department
: STREET: 9330 Zionsville Road
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: USA
: ZIP: 46268
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/036,987A
: FILING DATE: 09-MAR-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Stuart, Donald R
: REGISTRATION NUMBER: 28,479
: REFERENCE/DOCKET NUMBER: 50,608
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317)337-4816
: TELEFAX: (317)337-4847
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 80161 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-09-036-987A-1

Query Match      28.3%; Score 28.6; DB 3; Length 80161;
Best Local Similarity 64.2%; Pred. No. 13;
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY      34 GAGCCCCCGCTGCGAGGCGTGTGTCACGCGTGTGTCATGCTCCTCACTGGCG 93
      ||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 43852 GAGTGTCCGCTGACTGCGGTGTGTCACGCGCGCGGTCGCGAGCGCGACGTAGTG 43911
OY      94 GAGACGG 100
      ||||| ||
DB 43912 GAGATGG 43918

RESULT 13
US-09-370-700-1
: Sequence 1, Application US/09370700
: Patent No. 6274350
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard H.
: APPLICANT: Broughton, Mary C.
: APPLICANT: Crawford, Kathryn P.
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Treadway, Patti J

APPLICANT: Turner, Jan R
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 13987 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..13987
US-08-804-227C-13

APPLICANT: Turner, Jan R
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match      28.3%; Score 28.6; DB 4; Length 80161;
Best Local Similarity 64.2%; Pred. No. 13;
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY      34 GAGCCCCCGCTGCGAGGCGTGTGTCACGCGCTGCTCAGCTCATGCTCCTCACTGGCG 93
      ||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 43852 GAGTGTCCGCTGACTGCGGTGTGTCACGCGCGCGGTCGCGAGCGCGACGTAGTG 43911
OY      94 GAGACGG 100
      ||||| ||
DB 43912 GAGATGG 43918

RESULT 14
US-08-804-227C-13
: Sequence 13, Application US/08804227C
: Patent No. 3876991
: GENERAL INFORMATION:
: APPLICANT: DeHoff, Bradley S.
: APPLICANT: Kustoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 13987 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 350..13987
: US-08-804-227C-13
```

Query Match 28.1%; Score 28.4; DB 2; Length 13987;
Best Local Similarity 56.4%; Pred No. 13;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

OY 1 GTGGAACCGATGACAGCGCTGTTGCTGTCGAGACCCCGCTGCGAGGGGTGTGCAC 60
DB 12788 GCGTACGGAAGCCCTCCGCCGCTGCCGTCGACACCCCGCTGACCGGTGTGTGCAC 12847

OY 61 GCGCTGGCGCTCAGCGTCATGCTCCACTGGCGG 94
DB 12848 ACCGCCGCTGCTGGAAGCAGCGCCCTGACCG 12881

RESULT 15

US-09-372-422A-39/C
Sequence 39, Application US/09372422A

Patent No. 6313375

GENERAL INFORMATION:

APPLICANT: Rudolf Jung

APPLICANT: Francois Barrieu

TITLE OF INVENTION: Maize Aquaporins and Uses Thereof

FILE REFERENCE: 0919

CURRENT APPLICATION NUMBER: US/09/372,422A

CURRENT FILING DATE: 1999-08-11

PRIOR APPLICATION NUMBER: US 60/098,692

PRIOR FILING DATE: 1998-08-31

NUMBER OF SEQ ID NOS: 49

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 39

LENGTH: 1485

TYPE: DNA

ORGANISM: Zea mays

FEATURE:

NAME/KEY: CDS

LOCATION: (133)...(1015)

US-09-372-422A-39

Query Match 27.7%; Score 28; DB 4; Length 1485;

Best Local Similarity 60.5%; Pred No. 15;
Matches 46; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 3 CGAACGATGACAGCGCTGTTGCTGCGTCGAGCCCCCGCTGCGAGGGGTGTGCACGC 62
DB 882 CGTGCCGATGACGCGGCGCGAGGAATGATCCAGAGGCCCGTGAAGACGTTGCTGCCAC 823

OY 63 CGCTGGCGTCAAGCGTC 78
DB 822 CGCGGCGGCCAGCGTC 807

Search completed: November 5, 2002, 15:09:46
Job time : 214.737 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400

Perfect score: 101
Sequence: 1 gtcgaccgatgacagcgcgt.....cgtccactggtgcgagacgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	68750	21	AAZ55887
2	101	100.0	71989	21	AAZ55887
3	31.6	31.3	27541	22	AAZ29349
4	31.6	31.3	125401	22	AAZ17186
5	29.8	29.5	53789	19	AAZ21187
6	29.6	29.3	2787	23	AAZ52683
7	29.6	29.3	4851	22	AAZ52062
8	29.6	29.3	33529	23	AAZ17367
9	29	28.7	44377	18	AAZ78508

10	29	28.7	44377	18	AAZ80414	Platenolide synthase
11	28.8	28.5	3236	22	AAZ51552	Human polyomycet
12	28.8	28.5	9558	16	AAZ88228	Valencia orange ri
13	28.6	28.3	9513	22	AAZ88337	S. splinosa DNA fra
14	28.6	28.3	50000	22	AAZ88313	S. splinosa DNA fra
15	28.6	28.3	50000	22	AAZ88316	S. splinosa DNA fra
16	28.6	28.3	80161	20	AAZ21501	DNA fragment of Sa
17	28.4	28.1	13987	18	AAZ80415	Hybrid srng/cyG O
18	28.2	27.9	109519	22	AAZ08693	Micromonospora DNA
19	28	27.7	336	24	AAZ95178	Gene #1676 used to
20	28	27.7	336	24	AAZ65365	Lung cancer relate
21	28	27.7	1515	24	AAZ091596	M. capsulatus gene
22	28	27.7	28598	17	AAZ06769	Sorangium cellulos
23	28	27.7	28958	18	AAZ89956	Sorangium cellulos
24	28	27.7	28958	19	AAZ75289	DNA sequence of So
25	28	27.7	49377	21	AAZ05287	The sorapen biosy
26	28	27.7	88421	24	AAZ40781	88421nt genomic DN
27	28	27.7	4411529	22	AAZ199682	Mycobacterium tub
28	27.8	27.5	594	23	AAZ54048	Pseudomonas aerugi
29	27.8	27.5	1215	22	AAZ25116	Nucleotide sequen
30	27.8	27.5	2306	22	AAZ25117	Genomic DNA encodi
31	27.8	27.3	109519	22	AAZ08693	Micromonospora DNA
32	27.6	27.3	2532	24	AAZ90377	M. capsulatus gene
33	27.6	27.3	2532	24	AAZ72983	Methylococcus caps
34	27.6	27.3	3163	22	AAZ52536	Human polynucleoti
35	27.6	27.3	65140	22	AAZ17184	Streptomyces nous
36	27.4	27.1	1371	23	AAZ99933	Human alpha-2CAR v
37	27.4	27.1	1382	15	AAZ64890	Human derived adre
38	27.4	27.1	1383	23	AAZ9931	Human alpha-2CAR e
39	27.4	27.1	1869	22	AAZ59876	Human polynucleoti
40	27.4	27.1	1978	22	AAZ58090	Human polynucleoti
41	27.4	27.1	2084	24	AAZ35388	Human cDNA encodin
42	27.4	27.1	2742	23	AAZ51581	Pseudomonas aerugi
43	27.4	27.1	3126	22	AAZ94778	Human full-length
44	27.4	27.1	3948	24	AAZ98203	DNA encoding plako
45	27.4	27.1	3949	22	AAZ05388	Human caspase recr

ALIGNMENTS

RESULT 1	AAZ55887	standard; DNA: 68750 BP.
ID	AAZ55887	
XX	AAZ55887	
AC	AAZ55887	
XX	10-APR-2000	(first entry)
DE	Sorangium cellulosum	68.75 kb contig.
XX		
XX	Epithelone biosynthesis; type I polyketide synthase; taxol substitute;	
KW	anticancer; ds.	
XX		
OS	Sorangium cellulosum.	
XX		
PH	key	Location/Qualifiers
FT	CDS	1..1826
FT		/*tag- a
FT		/partial
FT		/product- "Partial Orf 1 protein (AAZ55880)"
FT		/note- "No initiation codon given in the specification"
FT	CDS	complement (1900..3171)
FT		/*tag- b
FT		/product- "Orf 2 protein (AAZ55881)"
FT	CDS	3415..5556
FT		/*tag- c
FT		/product- "Orf 3 protein (AAZ55882)"
FT	CDS	complement (5612..5992)
FT		/*tag- d
FT		/product- "Orf 4 protein (AAZ55883)"
FT	CDS	6226..6675
FT		/*tag- e

[illegible]

PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX	
PI	Schupp T., Ligon JM, Molnar I., Zinke R, Goerlach J, Cyr D;
XX	
DR	WPI: 2000-097741/08.
DR	P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
XX	AAY58592, AAY58593, AAY58594.
PT	New isolated epoethlone synthase genes, used for the recombinant
PT	production of epoethlone for use in cancer therapy -
XX	
PS	Claim 14; Page 87-104; 174pp; English.
XX	
CC	This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC	comprising 22 open reading frames (ORFs) and includes genes encoding
CC	proteins involved in the biosynthesis of epoethlones. Epoethlones A and
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC	starter unit; polyketides being synthesised from two-carbon building
CC	blocks, the beta-carbon of which always carries a keto group. Each round
CC	of two-carbon addition is carried out by a complex of enzymes known as
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC	EPoS A (AAY58573) and EPoS P (AAY58574) are involved in formation of
CC	the thiazole ring formation of epoethlones, and EPoS B, EPoS C, EPoS D
CC	and EPoS E (AAY58575-58578) are involved in polyketide backbone
CC	formation. EPo F (AAY58579) is an epoethlone macrolactone oxidase, and
CC	the proteins Orl 3 (AAY58582) and Orl14 (AAY58593) are thought to be
CC	involved in transport. Epoethlones mimic the biological activity of
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic
CC	compositions. Epoethlones exhibit a much lower drop in potency against a
CC	multiply drug-resistant cell line compared with taxol, and are
CC	considerably less efficiently exported from such cells by the multidrug
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC	epoethlones as anticancer agents, they are impractical to produce on a
CC	large scale. Epoethlones are too complex for industrial scale chemical
CC	synthesis and Sorangium cellulosum is difficult to ferment, producing
CC	poor yields of epoethlones. The nucleic acids of the invention may be
CC	used for the recombinant production of epoethlones in a heterologous host
CC	that is more amenable to fermentation.
XX	
SQ	Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
	Query Match 100.0%; Score 101; DB 21; Length 68750;
	Best Local Similarity 100.0%; Pred. No. 3,7e-18;
	Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 GTGCAACCGATGACACGCGCTTCGTCGTCGACGCCCGCCGACGAGGCTGTGCAC 60
Db	29912 GTGCACCGATGACACGCGCTTCGTCGTCGACGCCCGCCGACGAGGCTGTGCAC 29971
QY	61 GCCGCTGGCGTCACGCGTCATGCGTCACATGCGGAGACGCA 101
Db	29972 GCCGCTGGCGTCACGCGTCATGCGTCACATGCGGAGACGCA 30012
RESULT 2	
ID	AAA29349 standard; DNA; 71989 BP.
XX	AAA29349:
XX	12-SEP-2000 (first entry)
DE	Sorangium cellulosum epoethlone polyketide synthase operon genomic DNA.
XX	
KW	epoL: epoK: polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW	epoL: epoK: P450 epoxidase; ORA; ORB; promoter; enhancer; anti-fungal;
XX	tubulin polymerization assay; anti-tumour; cytostatic; ds.
OS	Sorangium cellulosum.
XX	
Key	Location/Qualifiers


```
FT CDS 3..992
FT /tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /tag= c
FT /label= epOA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5810
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6155
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493
FT /tag= h
FT /label= epOB_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature"
FT misc_RNA 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137
FT /tag= x
FT /label= epOC_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epOD_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574
FT /tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287
FT /tag= ap
```

```
FT      /note= "encodes ER5"
FT      30321..30869
FT      /tag= ag
FT      /note= "encodes KR5"
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      31440..32807
FT      /tag= as
FT      /note= "encodes K36"
FT      33018..34067
FT      /tag= at
FT      /note= "encodes A:6"
FT      34107..34676
FT      /tag= au
FT      /note= "encodes D16"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes E16"
FT      36705..37256
FT      /tag= aw
FT      /note= "encodes K16"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= epoc gene
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS1"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT1"
FT      41341..41922
FT      /tag= bb
FT      /note= "encodes KR1"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score: 101; DB 21; length 71989;
Best Local Similarity 100.0%; Pred. No. 3.7e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTCGACCGATGACGCGGTTTCTGCTGCTGACGCCCGCTGCGAGCGGTGTCAC 60
DB      24300 GTGACACCGATGACGCGGTTTCTGCTGCTGACGCCCGCTGCGAGCGGTGTCAC 24359

QY      61 GCGCGTGGCGCTGACGCTGATGCTGCACATG3CGGAGACGGA 101
DB      24360 GCGCGTGGCGCTGACGCTGATGCTGCACATG3CGGAGACGGA 24400

RESULT 3
ID      AAD17185 standard; DNA: 27541 BP.
XX      AAD17185;
XX      29-NOV-2001 (first entry)
DE      Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX      KW      Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX      anti-fungal; antibiotic; nys2; ds.
XX      OS      Streptomyces noursei.
XX      Key      Location/Qualifiers
FH      CDS      complement (454..1151)
FT      /tag= a "NysF protein"
FT      /product= "CDS does not include start codon"
```

```
FT      CDS      complement (1275..3092)
FT      /tag= b
FT      /product= "NysG protein"
FT      complement (3070..4824)
FT      /tag= c
FT      /note= "CDS does not include start codon"
FT      /product= "NysH protein"
FT      5122..6156
FT      /tag= d
FT      /product= "NysD3 protein"
FT      6338..27541
FT      /tag= e
FT      /product= "NysI partial protein"
FT      /note= "CDS does not include stop codon"

WO200159126-A2.
16-AUG-2001.
08-FEB-2001; 2001WO-GB00509.
XX      08-FEB-2000; 2000GB-0002840.
PR      10-APR-2000; 2000GB-0008786.
PR      14-APR-2000; 2000GB-0009387.
XX      (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA      (SNPE) SINTER STIFTETSEN IMD TEK FORSK.
PA      (ALPH-) ALPHARMA AS.
PA      (SINV-) SINVENT AS.
PA      (DZIE/) DZIEGEMSKA H.
PA      (ZOTC/) ZOTCHEV S B.
PA      (SEKU/) SEKUROVA O N.
PA      (EJAE/) EJAEVYK E.
PA      (BRAU/) BRAUTASET T.
PA      (STRO/) STROM A R.
XX      Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI      Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX      WPI: 2001-557614/62.
DR      P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX      New nystatin polyketide synthase polynucleotides and polypeptides,
PT      useful as antibiotics and antifungals -
XX      Claim 2: Page 151-166; 266pp; English.
XX      The present invention relates to the cloning and sequencing of the gene
CC      cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC      involved in the biosynthesis of the macrolide antibiotic nystatin.
CC      The nystatin PKS is useful as antifungal antibiotics. The present
CC      sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX      Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other:

Query Match      31.3%; Score 31.6; DB 22; length 27541;
Best Local Similarity 65.7%; Pred. No. 13;
Matches 46; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY      26 CGTCGTCGAGCCCGCGTGCAGGCGGTGTCGACGCCGTCGACGCTGATGCGTC 85
DB      13476 CGCTGTTGACACCGCGCGCTGCGGAGGTGTTGCCGAGACGCGACGAGCGCCG 13535

QY      86 CACTGCGGGA 95
DB      13536 CACTGCTGGA 13545

RESULT 4
ID      AAD17186 standard; DNA: 125401 BP.
XX      AAD17186;
```

XX 29-NOV-2001 (first entry)
 XX Streptomyces noursei nystatin PKS gene cluster DNA.
 DE Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KW antifungal; antibiotic; ds.
 XX Streptomyces noursei.
 OS
 FH Key Location/Qualifiers
 FT CDS 6337..34771
 FT /tag= a
 FT /product= "NysI complete protein"
 FT CDS 34792..51099
 FT /tag= b
 FT /product= "NysJ protein"
 FT CDS 51155..57355
 FT /tag= c
 FT /product= "NysK protein"
 FT CDS 57503..58687
 FT /tag= d
 FT /product= "NysL protein"
 FT CDS complement (58786..58980)
 FT /tag= e
 FT /product= "NysM protein"
 FT CDS complement (59045..60241)
 FT /note= "CDS does not include start codon"
 FT /tag= f
 FT /product= "NysN protein"
 FT CDS complement (60238..61296)
 FT /note= "CDS does not include start codon"
 FT /tag= g
 FT /product= "NysD2 complete protein"
 FT CDS 120628..121308
 FT /tag= h
 FT /product= "NysR4 (long) protein"
 XX MO200159126-A2.
 PN 16-AUG-2001.
 XX 08-FEB-2001; 2001WO-GB00509.
 XX 08-FEB-2000; 2000GB-0002840.
 PR 10-APR-2000; 2000GB-0008786.
 PR 14-APR-2000; 2000GB-0009387.
 XX (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIG.
 PA (SMTF) SINTEF STITTELSEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE/) DZIELEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (EJAE/) EJAERVIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 PI Zotchev SB, Sekurova ON, Ejaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX WPI: 2001-557614/62.
 DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
 DR AAE10149, AAE10150.
 XX New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX Claim 1; Page 188-254; 266pp; English.
 PS The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme

CC involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
 XX Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
 SQ Query Match 31.3%; Score 31.6; DB 22; Length 125401;
 Best Local Similarity 65.7%; Pred. No. 14;
 Matches 46; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
 QY 26 CGTCGTCGAGCCCCCGCTGCGAGGGGTGTGTCACGCCGCTGCGTCATGCGTC 85
 DB 13476 CGCTGTTGACCGCGCCCTGCGCGAGGTGTGTCGCCGAGAGCGACGAGCGCG 13535
 QY 86 CACTGCGCGA 95
 DB 13536 CACTGCTGGA 13545
 RESULT 5
 ID AAV21187 standard; DNA; 53789 BP.
 XX AAV21187;
 AC AAV21187;
 XX 24-JUL-1998 (first entry)
 DT Anycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
 DE Anycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
 KW polyketide synthase; actinomycete; ansamycin; ds.
 XX Anycolatopsis mediterranei.
 OS
 FH Key Location/Qualifiers
 FT CDS 1825..15543
 FT /tag= a
 FT /label= ORF_A
 FT /product= "polyketide synthase"
 FT CDS 13550..30759
 FT /tag= b
 FT /label= ORF_B
 FT /product= "polyketide synthase"
 FT CDS 30895..36060
 FT /tag= c
 FT /label= ORF_C
 FT /product= "polyketide synthase"
 FT CDS 36259..41325
 FT /tag= d
 FT /label= ORF_D
 FT /product= "polyketide synthase"
 FT CDS 41373..51614
 FT /tag= e
 FT /label= ORF_E
 FT /product= "polyketide synthase"
 FT CDS 51713..5293
 FT /tag= f
 FT /label= ORF_F
 FT /product= "polyketide synthase"
 XX MO9807868-A1.
 PN 26-FEB-1998.
 XX 18-AUG-1997; 97WO-EP04495.
 PR 20-AUG-1996; 96EP-0810551.
 XX (NOVS) NOVARTIS AG.
 PA Engel N, Schupp T, Toupet C;
 PI WPI: 1998-169172/15.

DR P-PSDB: AAW52845-W52850.
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
XX
PS Claim 4: Page 53-102; 205pp; English.

XX The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster DNA fragment from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polypeptides.
CC Syntheses, which can be used for assembling a library of polypeptides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

Query Match 29.5%; Score 29.8; DB 19; Length 53789;
Best Local Similarity 63.0%; Pred. No. 41;
Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 16 GCGCTGTTTCGTCGCGTACGCCCGCTTCGAGGTGGTGACGCCGCTGCGTCACG 75
DB 35428 GCGCTCTCTGTGCGCATTCAGCTTCGACCTTCGCGAGCTGCGGCCACCGCGCGCGC 35487

OY 76 GTTCATCGTCCAC 88

DB 35488 GGCACGTCGCCG 35500

RESULT 6
AA552683
ID AA552683 standard; DNA; 2787 BP.

AC AA552683;

DT 13-FEB-2002 (first entry)

DE E. coli DNA for cellular proliferation protein #405.

KW Antisense: ds: prokaryotic cellular proliferation gene;
antibiotic; antibacterial; drug des:gn.

OS Escherichia coli.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207127P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlson KL, Zyskind JW

Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR P-PSDB: AAU34824.

PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27: Seq ID No 6320; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 2787 BP; 734 A; 679 C; 769 G; 605 T; 0 other;

Query Match 29.3%; Score 29.6; DB 23; Length 2787;
Best Local Similarity 57.6%; Pred. No. 42;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

OY 2 TCGAACGATGACAGCGCTGTTCTGTCGTCAGACCCCGCTCGAGCGTGTGCACG 61

DB 275 TCGAACCTTTCGACGCGAGGTAAAGCATGGACTGCCCTGCTGTTCTGCGC 334

OY 62 CCGCTGCGCTCAGCGTCATCGTCACCTGCGC 93

DB 335 TAGAAGCGGACGACGATTATCGGTACTCTGCGC 366

RESULT 7
AAH52062
ID AAH52062 standard; DNA; 4851 BP.

AC AAH52062;

DT 04-SEP-2001 (first entry)

DE Mycobacterium tuberculosis potential drug target gene SEQ ID 116.

KW Drug target; growth; organism viability; characterisation; ds.

OS Mycobacterium tuberculosis.

PN WO200135317-A1.

PD 17-MAY-2001.

PF 13-NOV-2000; 2000WO-US31152.

PR 12-NOV-1999; 99US-0165086.

PR 12-NOV-1999; 99US-0165124.

PR 01-FEB-2000; 2000US-0179531.

PA (REGC) UNIV CALIFORNIA.

XX Eisenberg D, Rotstein SH, Marcotte EM;

DR	WPJ: 2001-329193/34.
DR	P-P8DB: AAG81211.
XX	
PT	Identifying nucleotide or polypeptide sequence for use as drug target,
PT	involves providing algorithm that analyzes a functional relationship
PT	between nucleotide or polypeptide sequences, and comparing the
PT	sequences -
XX	
PS	Disclosure: Page 131-133; 207pp; English.
XX	
CC	This invention relates to a method for identifying a nucleotide or
CC	polypeptide sequence that may be a drug target, or essential for growth
CC	or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC	represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC	tuberculosis proteins which are potential drug targets. The DNA and
CC	protein sequences are used to illustrate the method of the invention. The
CC	method involves providing an unknown nucleotide or polypeptide sequences,
CC	and comparing it to a number of sequences along with at least one
CC	algorithm capable of analysing a functional relationship between
CC	nucleotide and polypeptide sequences. The method is useful for
CC	characterising the function of nucleic acids and polypeptides that may be
CC	useful as a target for a drug or essential for the growth or viability of
CC	an organism.
CC	
SO	Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;
OY	
Query Match	29.3%; Score 29.6; DB 22; Length 4851;
Best Local Similarity	64.7%; Pred. No. 43;
Matches	44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Db	
9	GATGACAGCGCTGGTTTGTCTGCTGCATGAGCCCCCGCTGCGAGGAGTGTCACAGCCCTGG 68
I I I	I I
Db	3894 GTTGTGTTGCCGACGTGTGCGGAGTACC GCCGCGGTGCGGGGTATCATGCGCCGCG 3953
OY	
69	CGTCAGCG 76
I I I	I I I
Db	3954 CGTGCTCG 3961
RESULT 8	
AAS17367	
ID	AAS17367 standard; DNA; 33529 BP.
XX	
AC	AAS17367;
XX	
DT	12-MAR-2002 (first entry)
DE	
XX	DNA sequence of S. cellulosum polyketide synthase cosmid, PKOS28-26.
XX	
KX	Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW	acyl transferase domain; dehydratase domain; ketoreductase domain;
KW	acyl carrier protein domain; PKOS28-26; ds.
XX	
OS	Sorangium cellulosum.
XX	
PN	US6280999-B1.
XX	
PD	28-AUG-2001.
XX	
PF	31-AUG-1998; 980S-0144085.
XX	
PR	22-JAN-1998; 980S-0010809.
XX	
PA	(KOSA-) KOSAN BIOSCIENCE.
XX	
PI	Gustafson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX	
DR	WPJ: 2001-606536/69.
PT	Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT	polyketide open reading frame encoding modules with one or more domains
PT	such as ketosynthase, acyl transferase and acyl carrier protein domains

xx		Claim 4; Fig 1; 72pp; English.
ps		
xx		The present invention relates to the isolation of novel Sorangium
cc		cellulosum polyketide synthases (PKS), and the polynucleotide sequences
cc		encoding them. The polyketide synthases include catalytic domains such
cc		as ketosynthase domain, acyl transferase domain, dehydratase domain,
cc		ketoerductase domain and acyl carrier protein domain. A host cell
cc		comprising a PKS ORF (open reading frame) which encodes one or more
cc		more PKS domains is useful for producing polyketide synthases from which
cc		a library can be produced. The host cells are useful for constructing
CC		a colony with the ability to produce a particular PKS synthase and
CC		ultimately a particular polyketide. The polyketides produced by these
CC		colonies can be used collectively in a panel to represent a library or
CC		may be assessed individually for activity. Colonies in the library are
CC		also induced to produce the relevant synthases and thus to produce the
CC		relevant polyketides to obtain a library of candidate polyketides which
CC		can be screened for binding to desired targets such as receptors,
CC		signalling proteins, etc. The present sequence represents the DNA
CC		sequence of cosmid PKO828-26 which encodes one or more domains of
Cc		S. cellulossus PKS.
CC	Note:	The present sequence is said to encode the functional domains
CC		of S. cellulossus PKS which correspond to domains or domain subsets of
CC		the amino acid sequences of ORF1 (AAI10700) and OXF2 (AAI10701).
XX		
sq	Sequence	33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;
	Query Match	29.3%; Score 29.6; DB 23; Length 33529;
	Best Local Similarity	56.0%; Pred. NO. 46;
Matches	56; Conservative	0; Mismatches 44; Indels 0; Gaps 0;
OY	2	TGGAACCGATGTACAAGCAGCTGGTTCCTGCCTGACGCCCGCTGCGAAGGATGTGCACG 61 Db 24388 TCGCGGCGTGCTCGCGGAGGTGAGAGAAGTAAGACCCTGTGAAGGGGTTGGCGCAC 24447
OY	62	CCGTCGGCTCAAGCTCATGCTCCAACACTGCGGAGAACGA 101
Dd	24448	TGCGGCGCATAGGCCGCGCTGTTCGCCGAGAGCTCGA 24487
RESULT 9		
ID	AAT78508	standard; DNA; 44377 BP.
XX		
Ac	AAT78508;	
XX		
Dt	26-FEB-1998	(first entry)
XX		
DE		Platenolide synthase gene cluster.
XX		
KW		Platenolide synthase gene cluster; platenolide production; smg gene; multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX		
OS		Streptomyces ambofaciens.
XX		
FH	Key	Location/Qualifiers
FH	CDS	350..14002
FT	/tag=	-a
FT	/transl_except=	(pos:350..352, aa:Met)
FT	/note=	"ORF1 encodes protein shown in AAU23716"
FT		14046..20036
FT	/tag=	b
FT	/note=	"ORF2 encodes protein shown in AAU23717"
FT		20110..31284
FT	/tag=	c
FT	/transl_except=	(pos:20111..20113, aa:Met)
FT	/note=	"ORF3 encodes protein shown in AAU23718"
FT		31329..36071
FT	/tag=	d
FT	/note=	"ORF4 encodes protein shown in AAU23719"
FT		36155..41830
FT	/tag=	e

```
FT      /note= "ORF5 encodes protein shown in AAW23720"
XX
XX      EP791656-A2.
PN
XX      27-AUG-1997.
XX
XX      19-FEB-1997; 97EP-0301066.
XX
XX      22-FEB-1996; 96US-0012050.
XX
XX      (ELIL ) LILLY & CO ELI.
XX
XX      Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;
PI      Rosteck PR;
XX
XX      WPI: 1997-418047/39.
XX      P-PSDB: AAW23716-W23720.
XX
XX      DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT      for production of spiramycin-related polyketide antibiotics
XX
XX      Claim 9; Pages 8-33; 81pp; English.
XX
XX      This sequence represents the platenolide synthase gene cluster of the
CC      invention. This sequence is referred to as the srng gene, and was
CC      isolated from Streptomyces ambofaciens. This sequence encodes the
CC      multi-functional proteins which direct the synthesis of the polyketide
CC      platenolide. Platenolide is the basic building block of the macroide
CC      antibiotic spiramycin. The DNA can be used to produce compounds
CC      exhibiting antibiotic activity based on the platenolide structure,
CC      including specifically the macroide antibiotic spiramycin and spiramycin
CC      analogues and derivatives. Modifications of the platenolide synthase DNA
CC      sequence can be made so as to change the number and type of carboxylic
CC      acids incorporated into the growing polyketide chain and to change the
CC      kind of post-condensation process(s) that is conducted.
XX
XX      Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other:
SQ
Query Match      28.7%; Score 29; DB 18; Length 44377;
Best Local Similarity 55.4%; Pred. No. 67;
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY      1 GTCGACCGATGACAGCGCTGTTTCGTCTGTCGACGCCCGCGTGAGAGGGTGTCAC 60
DB      35820 GCGCGCGCCTGCTGTGAGAGCGGCTGTCCGCGACGAAACCGCGCGCGCGCGGAG 35879
QY      61 GCCGCTGGCGTCAAGCCTCATGCGCTCCACTGGCGGAGACCGA 101
DB      35880 CCGCGCGGGGGAGGCGCCGCGACCCACCGCCGCGACCGGA 35920
RESULT 10
AAAT80414
ID      AAAT80414 standard; DNA; 44377 BP.
XX
XX      AAT80414;
AC
XX      27-FEB-1998 (first entry)
XX
XX      Platenolide synthase gene cluster.
DE
XX      Ty lactone synthase gene cluster; tyjG gene; multifunctional protein;
KM      platenolide synthase gene cluster; platenolide production; srng gene;
KM      polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
XX
XX      Streptomyces ambofaciens.
OS
XX
XX      Key      Location/Qualifiers
FH      CDS      350..14002
FT      /tag= a
FT      /transl_except= (pos:350..352, aa:Met)
FT      /note= "ORF1 encodes protein shown in AAW22606"
FT      CDS      14046..20036
```

```
FT      /tag= b
FT      /note= "ORF2 encodes protein shown in AAW22607"
FT      CDS      20110..31284
FT      /tag= c
FT      /transl_except= (pos:20111..20113, aa:Met)
FT      /note= "ORF3 encodes protein shown in AAW22608"
FT      CDS      31329..36071
FT      /tag= d
FT      /note= "ORF4 encodes protein shown in AAW22609"
FT      CDS      36155..41830
FT      /tag= e
FT      /note= "ORF5 encodes protein shown in AAW22610"
XX
XX      EP791655-A2.
XX
XX      27-AUG-1997.
XX
XX      19-FEB-1997; 97EP-0301056.
XX
XX      22-FEB-1996; 96US-0012078.
XX
XX      (ELIL ) LILLY & CO ELI.
XX
XX      Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
PI      WPI: 1997-418046/39.
XX      P-PSDB: AAW22606-W22610.
XX
XX      DNA encoding Streptomyces fradiae ty lactone synthase domain - for
PT      production of tylosin-related polyketide compounds
XX
XX      Example 2; Pages 110-134; 220pp; English.
XX
XX      This sequence represents the platenolide synthase gene cluster of the
CC      invention. This sequence is referred to as the srng gene, and was
CC      isolated from Streptomyces ambofaciens. This sequence encodes the
CC      multi-functional proteins which direct the synthesis of the polyketide
CC      platenolide. Platenolide is the basic building block of the macroide
CC      antibiotic spiramycin. This sequence was used along with the tyjG gene
CC      (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tyjG
CC      gene is the ty lactone synthase gene cluster of the invention. The tyjG
CC      sequence was isolated from Streptomyces fradiae, and encodes
CC      multifunctional proteins which direct the synthesis of the polyketide
CC      ty lactone. Ty lactone is the basic building block of the antibiotic
CC      tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC      lacking the srng ORF1 sequence, or S. fradiae lacking the tyjG ORF1
CC      sequence, so that they can produce polyketides. The DNA sequence can be
CC      modified so as to alter the type of carboxylic acids incorporated, the
CC      number of carboxylic acids incorporated and/or the post-condensation
CC      reactions performed, thereby resulting in novel tylosin-related
CC      polyketides.
XX
XX      Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other:
SQ
Query Match      28.7%; Score 29; DB 18; Length 44377;
Best Local Similarity 55.4%; Pred. No. 67;
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY      1 GTCGACCGATGACAGCGCTGTTTCGTCTGTCGACGCCCGCGTGAGAGGGTGTCAC 60
DB      35820 GCGCGCGCCTGCTGTGAGAGCGGCTGTCCGCGACGAAACCGCGCGCGCGGAG 35879
QY      61 GCCGCTGGCGTCAAGCCTCATGCGCTCCACTGGCGGAGACCGA 101
DB      35880 CCGCGCGGGGGAGGCGCCGCGACCCACCGCCGCGACCGGA 35920
RESULT 11
AAK51552
ID      AAK51552 standard; CDNA; 3236 BP.
XX
XX      AAK51552;
AC
```

[illegible]

XX	AA088228;
AC	
XX	24-NOV-1995 (first entry)
DT	
XX	
DE	Valencia orange ribosomal RNA gene.
XX	
KW	Ribosomal RNA; rRNA; 18S; 28S; 5.8S; Valencia orange: RFLP analysis;
XX	DNA fingerprinting; restriction fragment length polymorphism; ds.
OS	Citrus sp.
XX	
FH	Key
FT	Location/Qualifiers
FT	/tag= a
FT	/product= 18S_rRNA
FT	1598..1848
FT	/tag= b
FT	/function= spacer
FT	1849..2011
FT	/tag= c
FT	/product= 5.8S_rRNA
FT	2012..2236
FT	/tag= d
FT	/function= spacer
FT	2239..5582
FT	/tag= e
FT	/product= 28S_rRNA
FT	5583..9396
FT	/tag= f
FT	/function= spacer
FT	7338..8291
FT	/tag= g
FT	/function= probe
FT	/note= "claimed"
FT	9397..9558
FT	/tag= h
FT	/product= 18S_rRNA
PN	JP07059577-A.
PD	07-MAR-1995.
XX	
EP	23-AUG-1993; 93JP-0227806.
PR	23-AUG-1993; 93JP-0227806.
XX	
PA	(KTIRI) KIRIN BEVERAGE KK.
XX	
DR	WP1; 1995-135901/18.
XX	
PT	Valencia orange ribosomal RNA gene - also probes and primers
PT	derived from citrus fruit RNA gene spacer regions, useful for
PT	RFLP analysis of citrus fruit
PS	Claim 1; Page 7-10; 18pp; Japanese.
XX	
CC	The ribosomal RNA gene has been isolated from Valencia oranges and
CC	sequenced (see AA088228). Primers based on sequences in the rRNA gene
CC	(see AA088228-Q88230 which are derived from nucleotides 1556-1575 and
CC	2257-2276 of AA088228, respectively) can be used to amplify regions of
CC	rRNA genes from various different types of citrus fruits. RFLP
CC	analysis and DNA fingerprinting of the amplified fragments allows
CC	different species to be distinguished.
SO	Sequence 9558 BP; 1853 A; 2675 C; 2996 G; 2034 T; 0 other;
Query Match	28.5%; Score 28.8; DB 16; Length 9558;
Best Local Similarity	65.6%; Pred. No. 72;
Matches 42; Conservative	0; Mismatches 22; Indels 0; Gaps 0;
9 GATGACAGCGCTGGTTTCGTGAGTGAAGCCCCGCTGCAGAGGTGTGCACGCCGCTGG	68

Db 7895 GCTGCGCGCATGTGCGCGCGCTGCGGCJCATGCTGCGCGCATGTGCTGCGCGCGCTGG 7954
 QY 69 CGTC 72
 Db 7955 CGCC 7958

RESULT 13

AAFB8337
 ID AAFB8337 standard; DNA; 9513 BP.

AC AAFB8337;
 XX

DT 28-AUG-2001 (first entry)
 XX

DE S. spinosa DNA fragment encoding ORF20, SEQ ID 45.
 XX

KM Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
 KM spinosyn; polyketide aglycone; transgenic plant; insect resistance;
 KM macrolide; insecticidal; polyketide synthase; ds.

OS Saccharopolyspora spinosa.
 XX

PN DE19957268-A1.
 XX

PD 08-MAR-2001.
 XX

PF 29-NOV-1999; 99DE-1057268.
 XX

PR 27-AUG-1999; 99DE-1040596.
 XX

PA (FARB) BAYER AG.
 XX

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
 XX

DR WPI: 2001-267102/28.
 DR P-PSDB; AAB70967.

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 PT recombinant production of insecticidal spinosyns and their derivatives
 XX

PS Claim 7a; Page 209-226; 354pp; German.
 XX

CC This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (I) to
 CC identify, inactivate or modulate genes involved in the biosynthesis of
 CC (II); (II) to generate a library of polyketide synthases; (III) for
 CC adding forosamine or trimethylrhannose to a spinosyn or polyketide
 CC aglycone; and (IV) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macrolides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence, ORF 20, encodes an S. spinosa polyketide synthase.
 CC
 XX

SO Sequence 9513 BP; 1392 A; 2759 C; 3637 G; 1725 T; 0 other;

Query Match 28.3%; Score 28.6; DB 22; Length 9513;
 Best Local Similarity 64.2%; Pred. No. 81;
 Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 34 GAGCCCCGCTGCGAGGCGTGTGACAGCGCGCTGCGTCACTGCTCCACTGGCG 93
 DB 8434 GAGTGTCCCTGACTGCGGTGTGTACACGCCGCGGGGTGCGCAGCGCGACGTAGTG 8493
 OY 94 GAGACGG 100
 |||||

Db 8494 GAGATGG 8500

RESULT 14

AAFB8313
 ID AAFB8313 standard; DNA; 50000 BP.

AC AAFB8313;
 XX

DT 28-AUG-2001 (first entry)
 XX

DE S. spinosa DNA fragment SEQ ID 2.
 XX

KM Forosamine; trimethylrhannose; polyketide synthase; biosynthesis;
 KM spinosyn; polyketide aglycone; transgenic plant; insect resistance;
 KM macrolide; insecticidal; ds.

OS Saccharopolyspora spinosa.
 XX

PN DE19957268-A1.
 XX

PD 08-MAR-2001.
 XX

PF 29-NOV-1999; 99DE-1057268.
 XX

PR 27-AUG-1999; 99DE-1040596.
 XX

PA (FARB) BAYER AG.
 XX

PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
 XX

DR WPI: 2001-267102/28.
 XX

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
 PT recombinant production of insecticidal spinosyns and their derivatives
 XX

PS Claim 7; Page 31-49; 354pp; German.
 XX

CC This invention describes a novel method nucleic acid (I) and its encoded
 CC polypeptide (II) containing at least one region that encodes an enzymatic
 CC activity involved in biosynthesis of spinosyns. (I) are used (I) to
 CC identify, inactivate or modulate genes involved in the biosynthesis of
 CC (II); (II) to generate a library of polyketide synthases; (III) for
 CC adding forosamine or trimethylrhannose to a spinosyn or polyketide
 CC aglycone; and (IV) for recombinant production of the corresponding
 CC enzymes, which are used for production of (II), their precursors or
 CC derivatives, including production of transgenic plants that express (II)
 CC and thus have increased resistance to insects. (I) are also useful as
 CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
 CC macrolides with insecticidal, but not antibacterial, activity, and can
 CC also be used to raise specific antibodies, useful for identifying
 CC expression clones in a gene bank. Cells transformed with (I) may produce
 CC (II) at significantly increased levels or produce new derivatives of
 CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
 CC genome which contains the coding regions for proteins involved in
 CC forosamine, trimethylrhannose and polyketide synthase biosynthesis.
 CC
 XX

SO Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 other;

Query Match 28.3%; Score 28.6; DB 22; Length 50000;
 Best Local Similarity 64.2%; Pred. No. 87;
 Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 34 GAGCCCCGCTGCGAGGCGTGTGACAGCGCGCTGCGTCACTGCTCCACTGGCG 93
 DB 18481 GAGTGTCCCTGACTGCGGTGTGTACACGCCGCGGGGTGCGCAGCGCGACGTAGTG 18540
 OY 94 GAGACGG 100
 |||||

Db 18541 GAGATGG 18547

RESULT 15
AAFB8316 ID AAFB8316 standard; DNA; 50000 BP.
XX AAFB8316;
XX 28-AUG-2001 (first entry)
DT S. spinosa DNA fragment SEQ ID 5.
DE
XX Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; ds.
OS Saccharopolyspora spinosa.
XX DEJ9957268-A1.
PN
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
PI Eberz G, Moehrl V, Froede R, Velten R, Salas JA;
DR WPI: 2001-267102/28.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PS
PS Claim 7; Page 74-91; 354pp; German.

This invention describes a novel method nucleic acid (I) and its encoded polypeptide (II) containing at least one region that encodes an enzymatic activity involved in biosynthesis of spinosyns. (I) are used (i) to identify, inactive or modulate genes involved in the biosynthesis of (II); (ii) to generate a library of polyketide synthases; (iii) for adding forosamine or trimethylrhamnose to a spinosyn or polyketide aglycone; and (iv) for recombinant production of the corresponding enzymes, which are used for production of (II), their precursors or derivatives, including production of transgenic plants that express (II) and thus have increased resistance to insects. (I) are also useful as markers for sequencing of the Saccharopolyspora spinosa genome. (II) are macroides with insecticidal, but not antibacterial, activity, and can also be used to raise specific antibodies, useful for identifying expression clones in a gene bank. Cells transformed with (I) may produce (II) at significantly increased levels or produce new derivatives of (II). This sequence represents a genomic DNA fragment of the S. spinosa genome which contains the coding regions for proteins involved in forosamine and trimethylrhamnose biosynthesis.

Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 other:

Query Match 28.3%; Score 28.6; DB 22; Length 50000;
Best Local Similarity 64.2%; Pred. No. 87;
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

34 GAGCCCCCGCTGCGAGGGGTGTGCACGCCGCTGCATGCGTCACATGCGC 93
||| |||||| | |||||| ||||| || ||| ||| | |
Db 22857 GAGTGTCCGCTACATCGCGTGTTACACGCCGCCGGGGTGTGCGAGAGCGACGATG 22916

94 GAGACGG 100
||||| ||
Db 22917 GAGATGG 22923

```
Search completed: November 5, 2002, 13:08:37
Job time : 246.215 secs
```


score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

(without alignments)
6406.881 Million cell updates/sec

Perfect score: 101

Sequence: 1 gtcgaaccgatgacagcgc...cgtccactggcgagagacgga 101

Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
```

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  genb1:.*
2:  gp_ba:.*
3:  gp_hlg:.*
4:  gp_in:.*
5:  gp_om:.*
6:  gp_ov:.*
7:  gp_pat:.*
8:  gp_ph:.*
9:  gp_pl:.*
10: gp_pr:.*
11: gp_ro:.*
12: gp_sy:.*
13: gp_un:.*
14: gp_yi:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_com:.*
21: em_ov:.*
22: em_ov:.*
23: em_pat:.*
24: em_pi:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_un:.*
29: em_yi:.*
30: em_hlg_hum:.*
31: em_hlg_inv:.*
32: em_hlg_other:.*
33: em_hlg_mus:.*
34: em_hlg_pln:.*
35: em_hlg_rod:.*
36: em_hlg_mam:.*
37: em_hlg_virt:.*
38: em_sy:.*
39: em_hltgo_hum:.*
40: em_hltgo_mus:.*
41: em_hltgo_other:.*

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF217189	AF217189 Sorangium
2	101	100.0	68750	1	AF210843	AF210843 Sorangium
3	101	100.0	68750	6	AR193029	AR193029 Sequence
4	101	100.0	68750	6	AR199551	AR199551 Sequence
5	101	100.0	68750	6	AR199559	AR199559 Sequence
6	101	100.0	68750	6	AR199567	AR199567 Sequence
7	101	100.0	68750	6	AR201097	AR201097 Sequence
8	101	100.0	68750	6	AR208671	AR208671 Sequence
9	101	100.0	71989	6	AR172664	AR172664 Sequence
10	67.4	66.7	10910	6	AX024483	AX024483 Sequence
11	67.4	66.7	10910	6	AX024476	AX024476 Sequence
12	40.2	39.8	36905	1	AE050965	AE050965 Streptomy
13	34.8	34.5	213050	1	AL646079	AL646079 Ralstonia
14	34.4	34.1	6286	1	AE009536	AE009536 Streptomy
15	34.4	34.1	34713	1	SCF42	AL371615 Streptomy
16	34	33.7	45396	1	SCF35	AL371610 Streptomy
17	32.8	32.5	1671	1	AX024451	AX024451 Sequence
18	32.8	32.5	1671	6	AX024444	AX024444 Sequence
19	32.8	32.5	2274	1	AX024442	AX024442 Sequence
20	32.8	32.5	2274	6	AE005686	AE005686 Caulobact
21	32.8	32.5	11992	1	AX024320	AX024320 Sequence
22	32.8	32.5	47713	1	AX024313	AX024313 Sequence
23	32.8	32.5	47713	6	AX024243	AX024243 Sequence
24	32.6	32.3	2030	14	SHE012474	AJ012474 Cercophit
25	32.4	32.1	137785	2	AP005065	AP005065 Oryza sat
26	32.4	32.1	176171	2	AP005067	AP005067 Oryza sat
27	32.2	31.9	30277	1	AY054120	AY054120 Mycobacte
28	32.2	31.9	104326	1	AB070940	AB070940 Streptomy
29	31.8	31.5	42816	1	SCC88	AL139288 Streptomy
30	31.6	31.3	27541	6	AX211706	AX211706 Sequence
31	31.6	31.3	123580	1	AF263912	AF263912 Streptomy
32	31.6	31.3	125401	6	AX211739	AX211739 Sequence
33	31.4	31.1	150039	2	AC114018	AC114018 Rattus no
34	31.4	30.7	169733	2	AC068046	AC068046 Homo sap1
35	31.4	30.7	12902	1	AE011870	AE011870 Xanthomon
36	31	30.7	134155	2	AP005517	AP005517 Oryza sat
37	31	30.7	142278	2	OS2N00135	AL662942 Oryza sat
38	31	30.7	148777	2	AE005410	AE005410 Oryza sat
39	31	30.7	167617	2	AC091403	AC091403 Oryza sat
40	30.8	30.5	3059	14	AF078729	AF078729 Cervid he
41	30.8	30.5	232941	2	AL670236	AL670236 Mus muscu
42	30.8	30.5	254197	2	AF389853	AF389853 Mus muscu
43	30.6	30.3	257717	2	AC105653	AC105653 Rattus no
44	30.4	30.1	146893	2	CNS08989A	AL732555 Oryza sat
45	30.2	29.9	2000	1	AB015764	AB015764 Halodacte

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
AF217189	AF217189	Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.	AF217189	AF217189.1	GI:7453554	Polyangium cellulosum. Polyangium cellulosum. Bacteria: Proteobacteria; delta subdivision: Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.	Myxococcales; Sorangineae; Polyangiaceae; Polyangium. 1 (bases 1 to 58733)	BCT 09-JUN-2000

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epoH1 gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epoH1 gene cluster from *Sorangium cellulosum*
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 10831849
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien, B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Korean Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source Location/Qualifiers
1..58733
/organism="Polyantrum cellulosum"
/strain="SMP4"
/db_xref="taxon:56"
1..992
/note="orf1"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62378.1"
/db_xref="GI:7453515"
/translation="VRGH/EAFADEGCVPRVLLYNLKNVVERHGDATIRFHTLLAL
SADYFEPRPVAIVAGNEIGVYERAIIVYREGFEARAVADGLNROKATWTSAL
DSWVEDRATYRAQFDDISVYLHRPDPDPDHERVEVEKTPYARPDLDYVPH
DTRRTVLADLSQVRLDNGQIVATVHRSMDGCOIEQPEHILRLVDEKREAREH
GIDRIRARASQARILYAEAGVNSAIARLLOLADGAAPELAEALYPERDTI
HIGAVROYDRRSERHLPVYSIVTIGENALVYVTHSTLTVALKDPD"
989..1501
/note="orf2"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62379.1"
/db_xref="GI:7453516"
/translation="MTDLPTETKRLKSLGFLGLACMEQLADKPMLEVALYPERE
HKSRLERIKNSRYAFAFPMTPDPSKIDRAVDLYDSRYADLFEVYPRYD
AKPLISTNKRAFADMGQYFPAACVTLVRLVRAVEYTEAESYRLKEKELNAT
RTKQRTTKH"
1501..1523
/gene="epoA"
1523..6263
/gene="epoA"
6263..6263
/gene="epoA"
6263..6263
/note="type 1 polyketide synthase; epoH1 gene cluster
module"
/codon_start=1
/transl_table=11
/product="epoA"
/protein_id="AAF62810.1"
/db_xref="GI:7453517"
/translation="MADRP::ERAADPIAIVAGCRLPGVITLSEFWTLLEGSDTV
GOVPEERLDAAMPDPLDAGKTPVTRASFSDVACDPAFSGISPEALMDPAHR
LLLEVCWEMDAEALAPSAIVGTETGVLGIPSEYEAALPATAEALDAGGTML
PSVAGRIISYVGLRGPCVIVDTAYSSSLVAVHLACVSLRSGESTALSLYALSP
STIVLAKTRALATDGRCKIPSAEADGRCGCAVVLKRLSGARAGDRLIAYIRG
SAINHDSGSLITVNGSSCELYLKRALADACAASSVGYEANTGTTLDPITQIA
LNAVIGLRDVAATPLIGSVKTLGHPETASGITGLKAVLSLQHGQIPAHHAQALN
PRISWGDRLTVTRATPWI DMNTPRAGVSSFGNSGNAHVLEEAAPACTPAPAE
RPAELLVISARTAAALDAHARLRDLETYSQCLGDAFSLATRSMEHRLVAAT
SSEGLRALDAAGQTPPGVVRGJADSSRGRLAFLFTGOGAQTGKMGRLVWPAF
REAFDLVRLNOELDRPLREYMAEPASVDALDLOTAFOTPLFTGEVYLAALMRS
MGVEPELVAGHSIGELVACVAGVSLDAVFLVARGRLMOLPAGGAMSTIAPEA
DVAAVAPHAASVSTIAVNGPDVYIAGAGPVHAAAMAARGAARTKALYSHAFHS
PLMAPLEAFGRVAVSSTYFRPSIYVNLGSKAGTDEVSFGIYVRAAREVYRAGD
VKALHAGAGTVEVGRKSTLDGLVPACLPDARPLALASRAGRDEPATVLEALGLM

AVGLVSNAGLEPPSGRRVPLPTYPQREERYWIDTKADDAAGDRRAGACHDEVERG
GAVRGDRSRARLDHPPESSGRRKVEVAAGRPPELEIDEGVLDRLVLYTERRAPG
LGEVEIADVAAGLSENDVOLALGAPDDLPKRPPLLEIGECAGRIYAVEGVNGLV
VGOPIVATLACGAFATHTTSAALVLPRLQALSAFPAAMPAYLTAWALGIRLOP
GERVLIHATAGGVGLAAYOMAGHVAEHAATAGPEKAREYISGVRVSDSRDPY
ADVRAWGGEGVDVNLISGELIDKSPNLRGFRVEIGKRCQYADPAAVNOIGLPERL
NLSFSLVDRGMLEPRAPVRLPEELIGLIRACVTFPPPIATPLPAAVNAFSSMO
NOHLKVLITLADPEVQIRIPTHAGAGSTGDRDLDAAPARAAALAEALRTQ
VSOVLRPEIRIVGAALFTRLGMDLMEYLRNRLEAKSLTSLTSTPLSTPINALLT
ONLDAITALSLSEVAENLRAGVDSFVSSGADWEIATL"
6260..10492
/gene="epoB"
10492..10492
/gene="epoB"
10492..10492
/note="non ribosomal peptide synthetase; epoH1 gene cluster
module 1"
/codon_start=1
/transl_table=11
/product="epoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453518"
/translation="MTINQLNELEHOGVKAADGERLOIAPKNAIPLNLAIRSEH
KSTILTMLRORLPAESIYAPAEERHVPPLDIQGSYVGLTGAFAVPSGJHAAREYD
CTDLIVARLSAERKVVARHMLRAHTLPDMQVTEPVVDADIEILDRIGDRSTREA
RLVSLRDAMSHRIYDTERPPLXHVAVRLDQCRRIYSLIDINVDGSLSTIKRWL
SFYEDPETSUYLELSTYRDYLALESRKSAHORSMDYWRRAVELPPPMLEKAD
PSTLEIRFRHEOMLPDSWSRLKORVGEKLTPTGVITLAFSEVIGRMSAPFTIL
NITTEIRLPHVRVNDITGDFSTVLIDITTRKDSFQRAKRIQEOLEWMDICDVS
GIEVQREARAVLGIQDGLFPVYLIDSLAQOQVGTSLQIRGTVPYVSTQPPOLLDH
OLYEHGDLVLAMVLDGVFPFPLDMLLEAVYVLRKLTPEEPSBOMRCSLPAPQLE
ARASANTSLNLEHTLGLFAVAEVLQPMOLAVASAKTLYEELSHRSRLARLR
BOGAPNTLVANVMEKMGEOVAVALVESAAVYIDADIPAEIRHILDLHGKVLV
LTPMLDCKLSMPGPGLIIVSDGVEDDGOGLPMPPIQVSDVAVYVTSGGVLPK
GMVIDHRCGAVNTIIDLNERFLEPGEDRALISLSPDSVDVGLIAAGCTIVVPA
SKTPRPAWMLIPEKRYVWNSVAPLAPMLVESEGPDSIARSLTSLSGWMPY
GILPGLQIRGVSTISLGGTAEASINSGITPVKRVDPMSASITRGRLRQOTVLD
EALERPVAWPEQLTIGGVGLALGTWRDEETKRSNVLVHPETGERLITGDLGRILPD
NIEFERGREDNQIKLNGRYVELGETLEELKSHPNRVDAIVPVGDAANKLLAVVYV
EGTRRRAEODASLKTERVADARAAKAKGLSDERVOFKLARGLRDLDGKVPVLD
TGLVREAGLDVYARRSRVRLLEAPIPFVEFRGLSCLSSVEPDGAALPFRKYPASG
STYPVQTYAVAKSGRIEGVDGSEFYVYHPEHRLKLSDHGIERGAHVQNDVDEEA
FGLLEFVGRIDALEISLSEKCLLEGGVAMLEQAPSCINVCVQGFDEOYR
PVLDRHSDVYVHGMIGRVDPROFOVTLQODSSPRATRTGAPPGDGHAFDILRD
FLRTKLPYVWTVFVELDALPITNSGVDRKALREKRDTSPPHSGHTRARDLEEL
LVAVYREVLGIEVGLQDLPYDICTATSHIYRMSLLOKRDRELATIELQYVNLGS
LASGRSDSKLEORPNQDYEARRKRRRS"
10639..16137
/gene="epoC"
16137..16137
/gene="epoC"
16137..16137
/note="type 1 polyketide synthase; epoH1 gene cluster
module 2"
/codon_start=1
/transl_table=11
/product="epoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453519"
/translation="MEQESSAIAVIGNSGRFPGARDIDEEFWNLROGTEAVQRFSEQ
ELASGVPAIVLPBNVYRAGSVLEVDVRFPAAGFISPRAEALMDPDHRTFMCAME
ALENAGYPTAEVGSIGVYAGANSSSYTLNHEPRAMPKRGFOULIGDKQYLAAT
HVSYLLNLRGSIISVQATCSLVAHSLACSLDLRECDMLAGITVIRIHRAGYVY
AAGGTFSPDGCAFEADKANGTIMGNGSVLLPLDPLADLGDGPVRAVILVAGVYD
GARKTGFTAPBEVNOQAALMALALAGEAASITVIEHNGTGLGATETGAALRVF
DRDASTRSCAIGSKYTGILHESAGIAGIKTYLALIEHQPLPSLSEPNSTIDF
NASPPYVNTSLKDMWTGSTPRRAGVSTGIGTNAHVYLEAPAKLPAAAPASAE
FVVSAAALDAALAAARLDHLOAHQGLSLDVAFLSATTSPMEHRLTAAAPASAE
REGDIAARGCTPGVAGRGSPCNVPPVVPFGOGSDOMGKROLAEDPVFHAL
SACBRALIAEGGMSLAEILADEESQERIVDDVOPVYFALAFAALALMRWGGVPA
VIGHSMGVAAVHAAGALSLDAVALICRRLRLRISGQEMAVTELSLAEAAALR
GGRPVSVAIVANSRSTVLSGEPALAGVSSILNAKGVGCRGVYDVVASHSPOYDPLR
EDILAAAGLGLPRAAIVPSTVYCAWAGCELTANVMNLRQVRAVYVQALOG
CHGLEFVMSPIPLITTYEERKRAAAGAAVGSIRAGQDRPMALELGLAQOGT
VPWGRLEFPAGGRVPLPTYPQREERYWIDTKADDAAGDRRAGACHDEVERG

Query Match	100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity	100.0%; Pred. No. 5e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 1	GTGCAACCGAGACAGCGCGCTTTCGTCGACGCCCGCGTCGCGGGCGTGTGCAC 60
DB 24300	GTGCAACCGAGACAGCGCGCTTTCGTCGACGCCCGCGTCGCGGGCGTGTGCAC 24359
OY 61	GCCGCTGGCGTCAGCGTCATCGTCCACTGCGCGAGACGGA 101
DB 24360	GCCGCTGGCGTCAGCGTCATCGTCCACTGCGCGAGACGGA 24400
RESULT 2	
AF210843	68750 bp DNA linear JCT 21-JAN-2000
LOCUS	
DEFINITION	Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
ACCESSION	AF210843
VERSION	AF210843
KEYWORDS	
SOURCE	AF210843.1 GI:6724237
ORGANISM	Polysangium cellulosum. Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangiineae; Polysangiaceae; Polysangium.
REFERENCE	1 (bases 1 to 68750) McInar,I., Schupp,T., Ono,M., Zirkle,R., Mliamov,M., Nowak-Thompson,B., Engel,N., Toupet,C., Strattman,A., Cyr,D.D., Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
AUTHORS	The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90
TITLE	

JOURNAL	Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE	20130945
PUBMED	10662695
REFERENCE	2 (bases 1 to 68750)
AUTHORS	Molnar, I.
TITLE	Direct Submission
JOURNAL	Submitted (03-DEC-1999) Natural Product Genetics, Novartis Agribusiness Research Institute, Inc., 3054 Cornwalls Rd, P.O. Box 12257, Research Triangle Park, NC 27709, USA
FEATURES	location/Qualifiers
source	1. .68750
	/organism="Polyanium celulosum"
	/strain="So ce90"
	/db_xref="taxon:56"
	<1. .1826
	/note="ORF1"
	/codon_start=3
	/transl_table=11
	/product="unknown"
	/protein_id="AAF26904.1"
	/db_xref="GI:6724238"
	/translation="ASDALFARATSARVLDGGRATERVHLAENGIEDLRALREH LRIOGSGPFHMCIGDLYVEILADHOPLASISPHNARSHPMTSDADVDPALV RMIAARGAPSPHCEERERARACQARLMAAPCPAPDIPREDANDIPGPP MSPEVAERLRASYATPELACALAMITGCGPSGTPAYEMDENILLCGLPT AIAASAGTSEALRGAARLFAWEVSSKKSQGLNIPALWERLFTIVRAGNADN LSRERAEALIAEVRRLAPAPAPAGAGLAVGSSGSLGIVDGLYSGDND IVMEGRIPSPVVLACTDPEFELAPLSOMLEFAHNAGTISKLESGPLVMARN OARMSIVHARGFMAYNOAMPDPERGAPFYVORSIMEFEHTPRLCHEPGSARS LACDEHLVNCESAGRIELMRHHRGAPSRAYIGENIATYPTSLTANHYL WADDRALIGVDKRTVEPVLIAETHPPAHVVEEDRPILALGDPDSDMVHERIR SGASTVADYQROKMDRPMVLANRGETFTNRLTLARS"
	complement(1900. .3171)
	/note="ORF2: similar to hypothetical proteins from Mycobacterium tuberculosis and Streptomyces coelicolor and to DD peptidases"
	/codon_start=1
	/transl_table=11
	/product="unknown"
	/protein_id="AAF26905.1"
	/db_xref="GI:6724239"
	/translation="MGALISVAPGCLGGAEEGQGDQAGALAPAREMAAEVA AGQKIDPVLVARGDVHVAVGTETELGSSAPMKRDTFRIASMTKAVTAVMMLE EGKIDGDPVDRMLPELANRKLVARIDGPIDETVPARPITVRDLMFTGFGISDPA SSPJORADIDELVNAQVPMTPHGPDEMTRLGTLELMHQGQOMYNGSLVQVLA VGRADGDFDAFYVERLRLAPLGMRDYDFHVPADKLAFAGCGYETDGTGKTDMDR GAEAYASPPAFPGSAGLYSTVDYILFRMLMNGVHGGRLLSAAYREMTADHL TPKQKASSTFPGEFETHGMYGMNAVTAAYSEVGRYGMGDSFTSWINDPRL IGVMTQASGFLTGALERFWSYVATESA"
	3415. .5556
	/note="ORF3: similar to Na/H and drug/H antiporters"
	/codon_start=1
	/transl_table=11
	/product="unknown"
	/protein_id="AAF26906.1"
	/db_xref="GI:6724240"
	/translation="MGCTERQYLVLTALITVTRASGELARLRLEVEIGLELG GVTVSGVGAALAPGHRALFOEPAVGVVLSLIGMIGALLMAGLEVQDILRKKA RPGALSIGAIAPPLIAGAPSAIYDRLPSGLFTGIVSVAYSAIVALLEREM RSTAOVTLAAGVSEYAAVLAAMTSSSTGSGBALVAANSALLASGLFLPMVYGR LTHLMMVADATKVSQGVSLVLTLEAALQRLGHLPLGAFLVGLVLSNPAT NRLPLDQVLTAVGLFAPVFFVLNMGKVDVDSOLTPAMQVTLALATARAAYVPA LGRRLGIGRSEALVAVGNMGKGTPLVAIVGELGLSNEAYTWYAVVALTVPA SVALLMLERAPTOESARLEREEARARAYIPGVRITVPAHLLPGFANDYES IVASKRLEGTVDITELSYEQAPGSPRAGSGLRARGRLRGCITRQREIRGS IQALIRSRDHLVIGARSPARAGSGERLQDAIVQRESNVLVYVGPAPARAS ARRLVPIGLEYSFAADLAHAVALAMDEVLSSAQDPPGAVVWRDEEPSVRV ARSLVEAIVGRRRLGVRSRVHVGAMPDEDTRELARAPYDILVLCGYDHGLGRL YLSSTVYVRSVYVALLVAHGTEQYR"
	complement(5612. .5992)
	/note="ORF4"
	/codon_start=1
	/transl_table=11

/product="unknown"
 /protein_id="AAF6907.1"
 /db_xref="GI:6724241"
 /translation="MDKFTGRCAIAEGYIPGSGNGPEOMTSFETACILNASDRDA
 QVATVITSDSDPAGPYRVTVPARTRHVRPNDTLEPEPIIPRODVASVIESDAPITV
 QHTRLDSROAENALLSTIAYNDRE"
 6226..6675
 /note="ORF5"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF2908.1"
 /db_xref="GI:6724142"
 /translation="MKHVDTRGRFRGRRIGHTLGLLASMALAGCGSPSEKTVQGTRLAP
 GADVAIVADVDPDAITLAVDVVHLSPPELLENGSEFFVWQPPSESPWRRCVIG
 VNADSRCKLAETTVVYNFELLITAEKSSPSSSAVIGPVSg"
 7610..11875
 /gene="epoa"
 7610..11875
 /note="EPOS A"
 /codon_start=1
 /transl_table=11
 /product="polyketide synthase"
 /protein_id="AAF26919.1"
 /db_xref="GI:6724253"
 /translation="MADHPITRAAEDPIATVYGASCRLPGGYIDLSGFHTLLESGRDV
 GRVPERMDAAMPDPDADGKTPVTRASLTSPVACPDASFEGRISPRALNPARH
 LLLEPCWALENNMAIAPESALVGTGTGVTGIGTGPETEALPQATASAEIDAHGCLGT
 PSVGRSTISALGLRGPCVAVDTITSSLVAVHILACSLRSGESTIALAGVSLMTSP
 STVLSLTRTALADGRC KAFSAADGREGRECAVVYLRKSGARDGOKILVING
 SAINDGSSGLTVPNGSDEIYLKRALADGCAASVGYEAGTGTGLDPIEIDA
 LNAVGLORDVATPLIGIVKTNLGHPEYVSGISGLTKVNLVLOHQIIPHLHQALNT
 PRISGDLRDVTRARTEVPDMNTPRAGVSSFGEMSTNAHVLEEPAPATCTPAPAE
 RPAAELVLSARTASALDQIARLRDLHETVYSGCLGVAFSLATTRSAMEHRLAVAT
 SREGRLALDAAAGOGTSZGAVRSIADSSRCKLFLFGOGAOTLGMGRGLDYVWSAF
 REAPDVCRLTNOELDRP..REVMAAFASVNDKLLDQTAFTOPALPTPEVALAMRS
 WCEPVEYAGSITSEIYACVAGVPSLEDVETLVANGRLQALPAGCAVSIAPRA
 DYAAVAVFHAASVSIAYVIAFDQVITAGQPVHAIAMAARGARTKALVSHAFRS
 PLMPALAEAFVAVESVTSRRPSITVLVNSLGRKACTDEVSSPGWVRHARVVFAGAG
 VKALHAAGAFVEVPEKSTLGLGVPACMPAPARALLASSAGDEBPATVLEALGGLM
 AVGGIVSWAGLEPSSGGRRVPLPTVPMQREERYIMDTKADDAARGDRAFGHDEVEEG
 GAVRGDGRSLRLDHPPEISGRREKVEAAGRPRLLEDGCVLILVYTEREAGG
 LGVEIYADGAFSTFNQYLGALGWPDLPKPNPMLLIGECAGRTVAAGEVNGVGL
 VGEVPIALNADGAFNHTVTSALVLPPOALSTEAMAPAYLTARYALDARLQOP
 GERLIHAATGSCGLIANYVQNHGAEVHAATGTPERKATLESIGVYVSDSRDREY
 ADVARNTGEEVDVYVNSI.SGELIDKSNLNRSHGRVELEGRDCTVADNOLGMPFR
 NLUSLFDRLRMMLERPAVVALLEELGLTAAAGVFPPTPIATPIARVADAFERSMOK
 AOHGKRLVLTIGDEPVOIHPTHAGAGSGTGRDLRLRASAAPARAANAALAEALRFO
 VSOVLRPEITVIGVAEALFTRLGMDSLNAVELENRNIEASLIKIKISTFELSTSPNIALLA
 ONLLDALITALSLERVAENLRAGVQNDVFSSGADQDMEITAL"
 7640..8920
 /gene="epoa"
 /note="Region: beta-ketoacyl synthase/malonyl-CoA
 decarboxylase"
 9236..10201
 /gene="epoa"
 /note="Region: acetyltransferase"
 10529..11428
 /gene="epoa"
 /note="Region: enoyl reductase"
 11549..11764
 /gene="epoa"
 /note="Region: acyl carrier protein"
 11872..16104
 /gene="epop"
 11872..16104
 /gene="epop"
 /note="EPOS P: contains condensation/heterocyclization,
 cytosine adenylation, putative NAD(P)H oxidase, and
 peptidyl carrier domains"
 /codon_start=1
 /transl_table=11

/product="nonribosomal peptide synthetase"
 /protein_id="AAP6925.1"
 /db_xref="GI:6724235"
 /translation="MTIINLLNEHQIKLADEGRILQIQAPKNALPNILARISEH
 KSTIIITMILRNPESIVPAPARNAAPFLPTDIOESYVWGRTGATVPSSGHAAVED
 CTDMDPLSTAFKFAVPRDMIRATLTDMMQVLEPKADIDILIDRGIDRSTRA
 RLVDLRMSIRSIITDEPRPLVAVRDLDEOTRPLVSLIDINDIGSLSIKDKL
 SPVEDPEISLPEVLEISROVLALSERKSESAHORSOMYKRIARLEAPPTLPKMD
 PSLKLEIFRTEDEPLPSDSMGLKARVREDSRLTPGYITLAAFSVIGRMSAPRFL
 NITLPLNRIPHAPRNDITGDFITSMVLLDITRDSFDQRAKIOEDLMEAMDCDS
 GIEVGEAKRPLVIGDAPFVUTLSALNQVAVGSLQRTGTPYVSTQPPQLDLHD
 QLEHEDGQLVLAVIDGAPFPPDLDDMLAAVAVLRLRTTEPMEQVRSCLPQAOLE
 ARASNAITNALISHTTHGFAFVAVLQLESGAAVYVARSKTLTLELSRSLRAKL
 BOGARNPTVALVMEKMEQVAVLVLESGAAVYVADMDPAEKVETILHDHGKVLK
 LTPWQGLCSMPGQIORTLISAGVGEQDQPPMKPISYDPSDLAAVYVYSGSGELK
 GYMDIHGKAVNTLIDINEREIGRQVRLALSLSPISDLYDPSDLAAGGTIVPDA
 SKLDPHAMEALLEREVATWNSVAPLMLALVEHEDGRDSLARKSLRLISGQVITP
 GILPELDAIRGVSVYLSIGGATGATSWISIGYPRVAVDLSAMSIPYGRPLRNQFTHVD
 BALPERVAPVQGLYISGGLVLAIGVRDEKRSFLVHPETGERLYTGDIGRYLFD
 GNIEMGEEDQILSKTERVAYLEILIEETLKSHPNRAVTVPGVMDANKLLAAYVP
 BGTRRAEADQSLKTERIDARNAHAAGVLDSDGEVOFKLARGLRDLDDKGVYD
 TGPDPRAEGVDAVRSRSTYGLLEAIPVEVGRGLCSIEYEDPDLKPEYPSAG
 STPYQVITAYKSRISLEGVDYEGTYYHPEHKLKLSDHGIRCAHRQNFVDEDA
 FNLLFEGVDAIESLYGSSRSRECLLEAAYQALLMEAPSCNTGCVCGQFNEEYQ
 PLTDLRHSYDVHQMGLGKRVDPQKGLGDDSPRRATVYKAPGGEHFAFMALDE
 FLRLKLEPMYVYVELDALDPLTSCYKDRALDEKRTDPSRSHGTHFADMLDEI
 LVAVRVELGLLEVYVGLQDSVFDGATSIHVYMRSLQKRLDREIATITELFOYPNLS
 LASGLRDSRLDQRPMDQORVVRKGRRS"
 16251..21749
 /gene="epob"
 16251..21749
 /gene="epob"
 /note="EPoS B"
 /codon_start=1
 /transl_table=11
 /product="polyketide synthase"
 /protein_id="AAP6920.1"
 /db_xref="GI:6724234"

	Query Match	100.0%	Score 101:	DB 1:	length 68750;
	Best Local Similarity	100.0%	Pred. No. 4,8e-14;		
	Matches 101;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GTGGAACGGAGACAGCGCGCTGTTTCGTCGCGGAGCCCGCGTCGAGGGGTGTGCAC	60		
Db	29912	GTGGAACCGAGACAGCGCGCTGTTTCGTCGCGGAGCCCGCGTCGAGGGGTGTGCAC	29971		
Oy	61	GCGCGTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA	101		
Db	29972	GCGCGTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA	30012		

RESULT 3	ARI93029	ARI93029	68750 bp	DNA	linear	PAT 20-APR-2002
LOCUS						
DEFINITION	Sequence 1 from patent US 6346404.					
ACCESSION	ARI93029					
VERSION	ARI93029.1	GI:20238994				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 68750)					
	Schupp,T., Ligon,J,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and					
	Gorlach,J.					
TITLE	Genes for the biosynthesis of epothilones					
JOURNAL	Patent: US 6346404-A,1 12-FEB-2002;					
FEATURES	Location/Qualifiers					
source	1..68750	/organism="unknown"				
BASE COUNT	9596 a	22436 c	25339 g	11159 t		
ORIGIN						

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 60
|||||
DB 29912 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 29971

QY 61 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 101
|||||
DB 29972 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 30012

RESULT 4
AR199551
LOCUS AR199551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 60
|||||
DB 29912 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 29971

QY 61 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 101
|||||
DB 29972 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 30012

RESULT 5
AR199559
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 60
|||||

|||||
DB 29912 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 29971

QY 61 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 101
|||||
DB 29972 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 30012

RESULT 6
AR199567
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 60
|||||
DB 29912 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 29971

QY 61 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 101
|||||
DB 29972 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 30012

RESULT 7
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 60
|||||
DB 29912 GTGGAACCATGACAGCGCTGTTTCGTGCGTGCAGACCCCGCTGCGAGGGGTGTGCAC 29971

QY 61 GCCGCTGGCGTCAAGCTCATGCGTCCATGCGGAGACCGGA 101
|||||

Db 29972 GCCCGTGGCTACGCGTATGCTCCATTTGGCGGAGACGGA 30012

RESULT 8
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zilkke,R., Cyr,D.,Dawn. and Gorlach,J.
TITL Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6383787-A 1 07-4AY-2002;
FEATURES
source 1..68750
location/Qualifiers
BASE COUNT 9596 a 22456 c 25539 y 11159 t

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACCGATGACAGCGCTGTTTCGTTGGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
Db 29912 GTCGACCGATGACAGCGCTGTTTCGTTGGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 29971

Qy 61 GCCCGTGGCTACGCGCTATGCTCCACTGGCGGAGACGGA 101
Db 29972 GCCCGTGGCTACGCGCTATGCTCCACTGGCGGAGACGGA 30012

RESULT 9
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 71989)
AUTHORS Jullien,B., Katz,L., Khosla,C., and Tang,L.
TITL Recombinant methods and materials for producing epoethliones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1..71989
location/Qualifiers
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 4.8e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACCGATGACAGCGCTGTTTCGTTGGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
Db 24300 GTCGACCGATGACAGCGCTGTTTCGTTGGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 24359

Qy 61 GCCCGTGGCTACGCGCTATGCTCCACTGGCGGAGACGGA 101
Db 24360 GCCCGTGGCTACGCGCTATGCTCCACTGGCGGAGACGGA 24400

RESULT 10
LOCUS AX024383 10910 bp DNA linear BCT 15-SEP-2000

DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024383
VERSION AX024383.1 GI:10184587
KEYWORDS
SOURCE Polyanquium cellulosum.
ORGANISM Polyanquium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangelineae; Polyanquilineae; Polyanquium.

REFERENCE 1 (bases 1 to 10910)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source 1..10910
location/Qualifiers
BASE COUNT 1852 a 3124 c 3613 g 2321 t

ORIGIN

Query Match 66.7%; Score 67.4; DB 1; Length 10910;
Best Local Similarity 79.2%; Pred. No. 3.5e-06;
Matches 80; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GTCGACCGATGACAGCGCTGTTTCGTTGGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
Db 1350 GCCGATCCATGACAGCGGCTGCTGGCCGCATCGAGCCCCCGTGGCGGGGTGTGCAC 1409

Qy 61 GCCCGTGGCTACGCGCTATGCTCCACTGGCGGAGACGGA 101
Db 1410 GCCCGTGGCTATGCTCCACTGGCGGAGACGGA 1450

RESULT 11
LOCUS AX024276 10910 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024276
VERSION AX024276.1 GI:10184550
KEYWORDS
SOURCE Polyanquium cellulosum.
ORGANISM Polyanquium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangelineae; Polyanquilineae; Polyanquium.

REFERENCE 1 (bases 1 to 10910)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source 1..10910
location/Qualifiers
BASE COUNT 1852 a 3124 c 3613 g 2321 t

ORIGIN

Query Match 66.7%; Score 67.4; DB 6; Length 10910;
Best Local Similarity 79.2%; Pred. No. 3.5e-06;
Matches 80; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GTCGACCGATGACAGCGCTGTTTCGTTGGTCGAGCCCCCGCTGCGAGGGGTGTGCAC 60
Db 1350 GCCGATCCATGACAGCGGCTGCTGGCCGCATCGAGCCCCCGTGGCGGGGTGTGCAC 1409

Qy 61 GCCCGTGGCTACGCGCTATGCTCCACTGGCGGAGACGGA 101
Db 1410 GCCCGTGGCTATGCTCCACTGGCGGAGACGGA 1450

RESULT 12
LOCUS AF509565 36905 bp DNA linear BCT 06-JUN-2002
DEFINITION Streptomyces sp. JP95 griseorhodin biosynthesis gene cluster,
complete sequence.
ACCESSION AF509565


```
VERSION      AF509565.1 GI:21039487
KEYWORDS
SOURCE      Streptomyces sp. Jp95.
ORGANISM    Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
REFERENCE   1 (bases 1 to 36905)
AUTHORS     Li, A. and Piel, J.
TITLE       A gene cluster from a marine Streptomyces encoding the biosynthesis
            of the aromatic spiroketal polyketide griseorhodin A
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 36905)
AUTHORS     Li, A. and Piel, J.
TITLE       Direct Submission
JOURNAL     Submitted (06-May-2002) Department of Bioorganics, Max-Planck
            Institute for Chemical Ecology, Winzler Strasse 10, Beutenberg
            Campus, Jena 07745, Germany
FEATURES
source      1..36905
            /organism="Streptomyces sp. Jp95"
            /strain="jp95"
            /isolate="gr1"
            /db_xref="taxon:195040"
            /clone="6C2"
            /cdate="2066"
            /gene="grh1"
            /protein="grh1"
            /codon_start=1
            /transl_table=11
            /product="putative transcriptional activator Grh1"
            /protein_id="AAM33679.1"
            /db_xref="GI:21039514"
            /translation="MDITSLVSENGITVPMTPAPRVQVLAIFAPAGVPIAY
            LTDELMAGRPPRSATLTQTYVLQRELITVALPAPGKATLAKDILVTCGGYLL
            NTRGTSIDVREERLGLAYRAGVGDGHAEGARLGSIALMRGAPFADVQIGOLEM
            EAARLESRLCALDORIEADRLGRHRELAELVYLSRYRTHENLRCOPMIALORSG
            RMSDLEVYRKRLGTLRSRELGLESLALRQGRSILLDHSQMTASDVPAVRAN"
            2296..3117
            /gene="grhr2"
            /protein="grhr2"
            /db_xref="GI:21039515"
            /codon_start=1
            /transl_table=11
            /product="putative transcriptional activator Grhr2"
            /protein_id="AAM33680.1"
            /db_xref="GI:21039515"
            /translation="MRIRVLGPLNADVNGSVSAGKROVLSTLALYPCQIMPVT
            LMEETGAPPVSAATLTQTYLILARRLGTAMGPEAPHTAKDYLATRHGGVLOIPA
            DSTDVHAERKVRREGQAFEAAGDDESADTLREALALMOGPALVDVRCGPALSTFVLR
            LEESKLVYVERRIDALRGHCEISALRDLITRAHPHNEGISHQAMALYLRSGRQAA
            ALDTRKLVRRIEELGVESPQOMRHOAMLSVDPALDLATGPRGSTFDLYAA"
            3315..4130
            /gene="grhd"
            /protein="grhd"
            /db_xref="GI:21039491"
            /codon_start=1
            /transl_table=11
            /product="putative thioesterase Grhd"
            /protein_id="AAM33656.1"
            /db_xref="GI:21039491"
            /translation="MAGTEPAVRILYCFAHAGAGDSVFHMAASNGEGEVPYLLPGR
            GGRREPRILTHDSLVNHLDRVRYRPAQGSALYGHSLGMAVAHAKALACALBEGILP
            SFVAVGCPPPDAASRLSDACESDSELEADLDRGALPEGVAAAGVYRRASLPLVLR
            DLRLAGALRAAARASAPRPLSVPLLAIVSGSRPLVSPATVGGWRMTSGPVAAKTV
            PGDHFVYRGEALPRLIGRACRVVRLTAGTGVNGRWPAETVFPIYIRPAVHP"
            4323..4583
            /gene="grhc"
            /protein="grhc"
            /db_xref="GI:21039492"
            /codon_start=1
            /transl_table=11
            /product="putative acyl carrier protein Grhc"
            /protein_id="AAM33655.1"
            /db_xref="GI:21039490"
            /translation="MQQLDLPITLIGLRTSGSGSPDHNLGNDLVNAPADLGYSLLCLL
            QVTGTERDAGVALDEQVLEAEATPRTQYIAVNVKLSQATA"
            4744..5715
            /gene="grhe"
            /protein="grhe"
            /db_xref="GI:21039492"
            /codon_start=1
            /transl_table=11
            /product="putative cyclase Grhe"
            /protein_id="AAM33657.1"
            /db_xref="GI:21039492"
            /translation="MPERAFHTSHRVDIAAPAGVYGLVADAVRMPLFPTVHVEA
            IDLBADRLRLNALACDDVTSWTRTLDRAARRIDPRQVPPSPBEGYGSWVEE
            LPEGSSRLTLRSALPASKATRTTPASWLERAHDDLALLDLKHLAERKATLDEHTVA
            VEETLRVDGPAELADFLYRIEMWPAQLPRLTRLDVEETPGIQGVRRDLADGGRH
            TARGVRCFPGGRITVYKIVPEPASTVQAGESWSEVPASGVTLRCRQAVLRCPDAP
            SDTEPRTESGGAGPEGRMLGRDSAVVLGLARRAAGSAVHTPKP"
            5777..6580
            /gene="grhf"
            /protein="grhf"
            /db_xref="GI:21039493"
            /codon_start=1
            /transl_table=11
            /product="putative phosphopantetheinyl transferase Grhf"
            /protein_id="AAM33658.1"
            /db_xref="GI:21039493"
            /translation="MPRLAPVHVHGGPDGPNNOHLHROWEDSGGVVLAITWGEMLPAVY
            TDPRRLGLIGDDWPEYRRRTADATVRRHRAARFVLYKHAATAALEDPALDLAARPG
            RPYLRGVGVGVSLAHTELGVVVGSRFTGRIGVDAEPRLTSLSVDVLLARACTPDERA
            AVKTLGGTGGAGAVTAAALLRLMTLKEATKGLGGLRPFAPFGVPAWBEAANAAYD
            EGLSPGSGPAPRAWTGARRALGRILVYARRHDAAPALRESSDRP"
            6920..8485
            /gene="grhg"
            /protein="grhg"
            /db_xref="GI:21039494"
            /codon_start=1
            /transl_table=11
            /product="putative transcarboxylase component of acyl-CoA
            carboxylase Grhg"
            /protein_id="AAM33659.1"
            /db_xref="GI:21039494"
            /translation="MTTETAASDPTAQLKQEAADGPDPEQATARQHAKGLTARE
            RIDLLDEGSGQEIEALRRHNASGSGLENKRPYTDGCVTTCGTYHGRRVRYANDERT
            FEGAGGEAHHAKIRKIDMALAAGAPLVSLNDGAGARKQESVTLAAGIGLFFRNKTLA
            SGVLPQISVMGLGPCAGGAAYSPALTDPLFVNRGTSOMKITGPDVVAKAVTGBEITONGI
            GADVAHAASGVAAFAHDEAACTEDVEFLSLSPNNRETAAPRAADPADRPDLDAL
            RDLVDANNOAYDVRAVTEIYDGEFEIOPMASNVICALTRLDGHTVGVANOPA
            AFAGVLDTHASEKARAFVQCDASIRPLVTLVDVPLGVDGHHGVIIRHGCARLTYA
            YCNATVPRVOLITRKAYGCAITWDSNSTGADVSLAMPFTNEIAYMGAGAAVYFRRE
            IAAADDPVAVRAQRKQYRTELMHPYIAERGLVDVYIDPATRSALIGALAMLRPKY
            TDLPSRKHNPPV"
            8502..8843
            /gene="grhh"
            /protein="grhh"
            /db_xref="GI:21039495"
            /codon_start=1
            /transl_table=11
            /product="Grhh"
            /protein_id="AAM33660.1"
            /db_xref="GI:21039495"
            /translation="MTPTTCAHYESAPNGTSGPHTERGNGTTEPPAAGCPAAPVPGI
            AISAAPARISGGSPTAETAAALAVLAVLIGRAAGGGRDAADAPGSTGTGMRPPRY
            RAPSGTWRGR"
            9055..9423
            /gene="grhi"
            /protein="grhi"
            /db_xref="GI:21039496"
            /codon_start=1
            /transl_table=11
            /product="Grhi"
            /protein_id="AAM33661.1"
```



```

predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE SENSOR TRANSDUCTION
HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17639.1"
/db_xref="GI:17430958"
/translation="MMKPASLAARLVLAFLVAGVMTGATVLYALNQLRATDD
AMLLGKIVQLRHCAKDQSIDIVDARRYEVSGHEGLILRVIGPGRFLVEANPE
GEPVAAAPAVPVDVLASESLATWMSRTGISRLIAAQRGNHEDVTVAQDGS
RAALMASYRRHIFLAVAGVAVALSLMLVSTLKGARIGROASVTVPOLDRLT
LEGAPRELSDLALNMLTRLQDGFARLSQFSPRLANLDFPTPLANGOTEVLTAF
RAGEYRAVLESLEEVARTSRMIEDMLFLARADHARVALSAPLANRHEIOVLYDT
EAVAADRITTOVCGNGVRADLTILRRAINLNDNLNRRHTPGKARIDVAOVLPST
AIAVRNPGGIPASLPLFLDFRFYRGDPANRSERAGLAIIVQITMDLHGTAQAS
SPDGLTFEFLIFPA"
/complement(3145..3819)
/gene="czcR"
/feature="RSP0489: RS00367"
/complement(3145..3819)
/gene="czcR"
/function="cell processes; protection responses"
/feature="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17640.1"
/db_xref="GI:17430953"
/translation="MRLIVDEDEPKAGDYLKGTESGFVADLARSGPDGLYQAMEHD
YDLIVDMLPGMDGMQVIRELRKRKTPTVLFIARDELSDRIGLEAGADYAKPF
AFELVARVHTILRGPMRESEVLEIADLTIDTIRKRTTRAGRRIDLTAKEXALIHLL
AKRTGEVLSRLISSQVMDVNFDSNTNVDAIRLRAKVDPPFEHKLITLRLGKGY
LDIVAR"
/4089..4394
/gene="RSP0490"
/feature="RSP0368"
/4089..4394
/gene="RSP0490"
/function="miscellaneous; hypothetical/global homology"
/feature="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD17641.1"
/db_xref="GI:17430960"
/translation="MSVRNCLILPLAAAAFLACGAATAASVSEQALSPLOHGARDP
YTDAGRTRDPTTGGAQAEATLQASVGRDPYSDGMSRSSAPRDRDPYTDGAQA"
/4400..5725
/gene="czcC"
/feature="RSP0491: RS00369"
/4400..5725
/gene="czcC"
/function="cell processes; protection responses"
/feature="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
RESISTANCE PROTEIN"
/protein_id="CAD17642.1"
/db_xref="GI:17430961"
/translation="MWHPCIRAAVVALSCALLVBSGLAAPAGAAASPTAPAAQS
APASVGAHEHMTLQOALIDRALATHPALAAGYAVRASEGAVQASRLRPEAAFLT
EDVRDRATMTAQNIPLEIGKRRARTRAELRDAAGOPAAVLAELRADYRRAF
ELSTAOERLTLAQSADLAAQAEAAVVRVQGVSPYEQKAAVARAGATLDRDQ
ADLYAQRALSAFNGAVAVPEADGSAETLEFAYPEEARVYLAATAEPVALASTPPA
ILRARLEIRQALNVERSKVPVYTLVSQAKRDSGANANNAVIGIAPLEFDRNG
SILEROALDAADYERLRLLEQAAALSQDVARLEAQAAGALRODVLPGARAYDA
ARIFFEAGKFNFLVDLPAQRITLFOARQYLAALSRAHQAAAGIDRIILGR"
/5752..6966
/gene="czcB"
/feature="RSP0492: RS00370"
/5752..6966
/gene="czcB"
/function="cell processes; protection responses"
/feature="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE COBALT-ZINC-CADMIUM RESISTANCE (CATION
EFFLUX SYSTEM PROTEIN) TRANSMEMBRANE"
/protein_id="CAD17643.1"
/db_xref="GI:17430962"
/translation="MTLPROOQRAVIAITVALGALAGALTMWGARORAAHSEAEOP
AAHAEHQAPKAEAAEAPHAETIMTAEATANALIGVTTAGAAQLRFPVLPGLIRN
DDRRAHVPRVAGVAYVELGQPYTRGOLAYLSSALSDQSELOAQQRLALAO
EAAHEKRLMEQGLAAQDYQARLAEIADRTADNARQIAAIGAPAGTAVFEL
RAPDGVVVAHLSQGEAVQAEAAVFIADRTWADPAVYAKLEAVTTATVTRA
TATGAVQGVKSYGALLGEQTSAPAPVATIDNKRAMPMPFVSYSVSGSRVVPVA
VRAVAVQVDEKPSVFVPPVAFVFRPKYTRSGQTMTEIIDGLPAGTRVAANSYL
KAEGKSNADHH"
/6977..10171
/gene="czcA"
/feature="RSP0493: RS00371"
/6977..10171
/gene="czcA"
/function="cell processes; transport of small molecules"
/feature="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
Query Match 34.5% Score 34.8; DB 1; Length 213050;
Best Local Similarity 60.6%; Pred. No. 76;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 7 CCGATGACAGCCGCTGTTCTGCTGCGAGCCCGCTGCGAGGGGTGTCGACGCCGCT 66
DB 9923 CTGATGACGGCGCTGTTCCCTCGCTCGGCTTCGTCGCGATGGCGTGGCAGCGGACC 9982
QY 67 GCGCTCAGCGTCATGCGTCCACTGCGCGAGACGG 100
DB 9983 GCGCGCAGAGTCACGCTCCGCTGCGACGGTGG 10016
RESULT 14
AF009336 6286 bp DNA linear BCT 06-OCT-1999
LOCUS
DEFINITION Streptomyces coelicolor lipase (lipA), and lipA activator (lipR)
ACCESSION AF009336
VERSION AF009336.1 GI:4102168
KEYWORDS
SOURCE
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
```

REFERENCE 1 (bases 1 to 6286)
AUTHORS Valdez,F., Gonzalez-Ceron,G., Kleser,H.M. and Servin-Gonzalez,L.
TITLE The Streptomyces coelicolor A3(2) lipA operon encodes an
JOURNAL extracellular lipase and a new type of transcriptional regulator
MEDLINE 99445177
PUBMED 10517589
REFERENCE 2 (bases 1 to 6286)
AUTHORS Valdez,F. and Servin-Gonzalez,L.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-1997) Molecular Biology, Biomedical Research
Institute, National Autonomous University of Mexico, Apartado
Postal 70228, Ciudad Universitaria, D.F. 04510, Mexico
FEATURES
source 1..6286
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2) M145"
/db_xref="taxon:110226"
complement(1..514)
/note="similar to the E. coli MCA protein"
CDS
/codon_start=1
/transl_table=11
/product="orf1"
/protein_id="A009314.1"
/db_xref="GI:4102169"
/translation="MRICLGTGPMADMAHAPALAGHDDLDFAGVWGRPDAAEELAO
RHGTBAVDLADLADYLAAYALPPAYAGLAVARAAGCHLLDKPLASTYAGCRA
YVAVYATADVASYVFTTTRQPTETAMITRQSSRDHFTARAKMLCAVFGEDSPAD
SPWRKRGALMDYGFHALSVLPFLGDAKRVAAA"
975..1907
/gene="lipA"
975..1907
/gene="lipA"
/note="extracellular lipase"
CDS
/codon_start=1
/transl_table=11
/product="lipase"
/protein_id="A009315.1"
/db_xref="GI:4102170"
/translation="MOONHTHAAPGARPVLRGVRRLAAVTAAYAVLVLTGLGP
GAQADNPYRGAPTES;IEALRGYSVADSVSLAYEGGCGTITPTSDGTG
GAVYAPGTAAYOSTAMLCPLASOGVAFVFTDTMTTLDQPSRROLAALDYTG
RSSVGRIDSGRLGVNGH;KGGCGTLEAKSRSLQALPLTWNLDKMPYSTPL
VWGADDTTAPVASHAEPYVSLGPSSTDAVELNATHPSPTMTTAKYSISWLK
RFIDDDTRQFQCPRLPSLTIEYRGNCPHGS"
2078..4903
/gene="lipR"
2078..4903
/gene="lipR"
/note="putative transcriptional activator of lipA gene"
CDS
/codon_start=1
/transl_table=11
/product="lipR activator"
/protein_id="A009316.1"
/db_xref="GI:4102171"
/translation="MTGOTFVTRAVRVHRSNORALRAMTALPAHGRLIAGSPG
LGRITLQMAASRFAGP;LHGCPGDNAHILACGLLTIRAAVAAVYLVVDAAH
RMDAPRTELARARLSTADRYGLLSVASDRAVPEFRALPVVHDLPAKRAAL
LDDLTDGVAAPAREQLA;AEGNPALLVALVRLSPAEIRGRPLPSLAEGVLA
AAGVPTGSAHARDLLT;AAAGSDDSDADAGLVADLRRLRPAPOHPIGIDP
GIGDPGAPPLPOLVLAGIRLFRSPILVRVATAAADRRAVRAALADLDGGH
RLTALHRSMSLDGPAPAPISADRLATAADGVPASHRLRATVYARAELVADGAR
AERYTAAEOLLAGRPRRLPLAAARDAPADVVRGRALLRGWTELDDGPVGAH
OSULLAASILADDDPRAALAAADAAAMAGDPLACIATLAPRPLAERHSPSD
GPAARTAAVVDHHRVGMRLLEGRLDRAGAPLGQVVRSGTGDREPTILKSAALL
LGDVDAARGGAALAAARKLGSALVPALELVAELRAGRHDPARTHAEGSLTA
LRAGONTAAHRAVLAALF;SIEEPDAVARHYAALDTRRRRLAQAAALQMAAR
ADLGRPRPDALRGLVLI;PGRGHFAVRHYALPCEFAVLAALAGDADARADAR
ADWAARGDPOAAALARCHALLAPAEHADLYRRALDRHDEAGDFFERAPARILYK
WLRRRRPREARGRLGTALGFERGAGYMAQARDELALGAPSPAGACTPARL
GRITPODLIAROVAGATNREVALTLGSTRIVDHLKRVFALGVRSLTELARWG
QARAGDAQADARRAGHCKERTGAP"
5024..5956

/note="similar to protein encoded by gene downstream of
the Streptomyces exfoliatus M11 lipR gene"
/codon_start=1
/transl_table=11
/product="orf4"
/protein_id="A009317.1"
/db_xref="GI:4102172"
/translation="MOHAYRSRAAIRTGILTPVRPRRTGVTSLIDADALRVLHRAARA
LDDLDPLDTRMALIREQEPAYRAALTKDPTAQEATAPCGTACSLDPRGARDAA
RRCSMRIGARAEOGLPLDALHAFRLGSLVMORLEETSRADAEVRLVHVADY
WNPVDEHCFVLADYAROTEMOGLRRRENVRDLAIGDTRIDLEAARALGLE
HRRVYVAACGPAPAGPTPRRRTGSGTGTGCTGCTGCGGACGAGCAGCACC 3796
RRRTSPSPASGSVSAARSTDSPLSGTPVGSRRTP"
BASE COUNT 745 a 2625 c 2239 g 677 t
ORIGIN
Query Match 34.1%; Score 34.4; DB 1; Length 6286;
Best Local Similarity 60.9%; Pred. No. 1.5e+02;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 7 CCAGTACAGCGCTGCTTTCGTCGTGACGCCCGCTGCGAGGCTGTGCACGCCGT 66
111 1111 1111 11 111 1 1 11111 111 11 1111 11
Db 3737 CCGAGACACTGCTGCGGTGCGGCGCGCTGCTGCTGCGGACGAGCAGCACC 3796
11 1111 11 11111 111 11
QY 67 GCGTCAGCGTCATGCGTCCACTGCGGAGAC 98
11 1111 11 11111 111 11
Db 3797 CGCAGCGCGCGCGCGCGCGCGCTGCGCGCGC 3828
RESULT 15
SCF42/c 34713 bp DNA linear BCT 12-MAY-2002
LOCUS Streptomyces coelicolor cosmid f42.
DEFINITION AL137165 AL645882
ACCESSION AL137165.2 GI:20520910
VERSION AL137165.2
KEYWORDS ABC transporter protein, ATP-binding component, Atrac-family
regulator; branched-chain amino acid ABC transport permease;
branched-chain amino acid transport ATP-binding protein;
DNA-binding protein; glycosyl hydrolase; glycosyl transferase;
hydrolase; integral membrane protein; lipA; lipase; lipoprotein;
lipR; membrane protein; oxidoreductase; transcriptional activator.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
REFERENCE 1 (bases 1 to 34713)
AUTHORS Actinomycetaceae; Streptomycetaceae; Streptomyces.
Kinsahl,H., Kleser,H.M., Denapate,D., Elchner,A., Cullum,J.,
Redenbach,M., and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
JOURNAL for the 8 Mb Streptomyces coelicolor A3(2) chromosome
MEDLINE Mol. Microbiol. 21 (1), 77-96 (1996)
PUBMED 8843436
REFERENCE 2 (bases 1 to 34713)
AUTHORS Oliver,K. and Harris,D.
JOURNAL unpublished
REFERENCE 3 (bases 1 to 34713)
AUTHORS Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT On May 9, 2002 this sequence version replaced g1:6752308.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
details of S. coelicolor sequencing at the Sanger Centre are
available on the world wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are
numbered using the following system eg SC787.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
5024..5956

	RESULT 1		
CNS04HMV/c			
LOCUS	CNS04HMV	630 bp	DNA linear GSS 21-MAY-200
DEFINITION	Tetradodon nigriviridis genome survey sequence PUC-ori end of clone 110F12 of library G from Tetradodon nigriviridis; genomic survey sequence.		
ACCESSION	AL291136		
VERSION	AL291136.1 GI:8029716		
KEYWORDS	GSS; genome survey sequence.		
SOURCE	Tetradodon nigriviridis.		
ORGANISM	Tetradodon nigriviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acartinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.		
REFERENCE	1 (bases 1 to 630)		
AUTHORS	Roeest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winczer,P., Brottier,P., Quettler,F., Saurin,W. and Weissenbach,J.		
TITLE	Human gene number estimate provided by genome wide analysis using Tetradodon nigriviridis DNA sequence		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 630)		
AUTHORS	Roeest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quettler,F., Saurin,W., Bernot,A. and Weissenbach,J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigriviridis		
JOURNAL	Unpublished		

REFERENCE	3 (bases 1 to 630)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000)
COMMENT	This sequence is a single read and was generated as part of a larger

<http://www.gemscope.chs.ch/1ec1ad0011>

source	1. .630
--------	---------

/note="Genoscope sequence ID : CUBG110DC06SP1-end : PUC-Or1"

BASE COUNT	121 a	198 c	113 g	158 t	40 others
ORIGIN					

Query Match	40.2%	Score 40.6	DB 17	Length 630
Best Local Similarity	57.1%	Pred. No. 8.6		
Matches 52	Conservative 10	Mismatches 29	Indels 0	Gaps 0

```

OY      5  GGTTCGATCGCGCGCGTCAATGGGCGG-AGCAGGTGTCATCGCGGGCGTGGAGCAAG 64
          || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     492  GCGCGAGGTTCCGCGCGCGGCGGTGCGCC'TGAGAGTAAATAAARKMGTGKTGGCGGAGGC 433

```

Qy	65	G T G C A G G C G A T C G C G C G G G G T T C G C G G C : G	95
Db	432	G G T G G A G G T G G T G G C G C G G C G T G C C G C : G	4022

RESULT 2	LOCUS	DEFINITION
BF588146	47 bp	linear EST 12-DEC-2000
BF588148	FM1_38.C12.b1.A003	Floral 1-induced Meristem 1 (FM1) Sorghum
	proteiquum cdna, mRNA sequence.	

ACCESSION	BF588148
VERSION	BF588148.1
KEYWORDS	EST.
SOURCE	Sorghum proflinquinum.
ORGANISM	Sorghum proflinquinum

REFERENCE
1 (bases 1 to 470)
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

TITLE	JOURNAL	COMMENT
An EST database from Sorghum: floral-induced meristems unpublished (2000)		Contact: Cordonnier-Pratt MM

Email: mmp@pratt.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

FEATURES	Location/Qualifiers
source	1. .470

```

/organism="Sorghum prostratum"
/db_xref="taxon:132711"
/collection.lib="Floral-Induced Meristem 1 (FM1)"
/note="Organ: Floral-Induced meristems; Vect

```

plusvescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being

BASE COUNT 63 a 151 c 187 g 69 t
ORIGIN

Query Match	39.68;	Score 40;	DB 12;	Length 470;
Best Local Similarity	63.58;	Pred. No. 12;		
Matches 61; Conservative	0;	Mismatches 35;	Indels 0;	Gaps 0;

QY 5 GGTTCGATCCGGCGGCCTCAATTGGCCGAGACAGTGTGATCCGGCGCCTTGAGACAAGC 64
 ||||| || | |||| | | ||| ||||| | | | |||
Dd 224 GGTCGCCGTACCAGTCTGTCGTGCGCGTCTCTGTTGCTGATGTCAGAGTTCGA 283

QY	65	GGTGCAGCGCGATCCGCGCGGGTTCGCGCGCGCG	100
Db	284	GGTGCAGCGCGGTGCGCGCGCGCGCGGAGCATTG	319

RESULT 3					
B0621957					
LOCUS					
DEFINITION					
B0621957	560 bp	mRNA	linear	EST 01-JUL-2002	
cdic.pK002.18	Candidiobolus cornatus ARSEF	512	Candidiobolus		
cornatus cDNA,	mRNA sequence.				

ACCESSION	B0621937
VERSION	B0621957.1
KEYWORDS	EST.
SOURCE	Condlobolus coronatus.
ORGANISM	Condlobolus coronatus

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 560)	Freimoser, F. M., Screen, S., Hu, G. and St. Leger, R. J.	EST analysis of genes expressed by the zymocorn pathogen <i>Conidiobolus coronatus</i> during optimized secretion of proteins	unpublished (2002)	Contact: Freimoser F. M.

UNIVERSITY OF MARYLAND
4112 Plant Sciences Building, College Park, MD 20742, USA
Tel: 301 405 16 13
Fax: 301 314 92 90
Email: ff34@umail.umd.edu.

FEATURES
Location/Qualifiers
source
1..560

BASE COUNT	ORIGIN
103 a	110 c 181 g 145 t 21 others

Query Match	39.28;	Score 39.6;	DB 14;	Length 560;
Best Local Similarly	63.88;	Pred. No. 14;		
Matches 60; Conservative	0;	Mismatches 34;	Indels 0;	Gaps 0

qy 1 CGTGGGTGTCGATCGCGGCGGTCATGGGCGGAGCAGGGGTGATCGCGGGCGTGAGC 60
|| || | ||||| |||| | || || ||||| | || |||||
Db 51 CGGTGAGGTGATCGCGGTGGTCGCGGAGGCGGTCGTGGGTCTTCAAGGCTGAGG 110

```
QY      61  AAGCGTGACGGCATCGCGCGGGGTTGCGCGC  94
          ||||| ||| ||||| ||||| ||| |||
Db     111  TCGCGTGAGAGTATCGCGCGGCTCTGAGAGC  144
```

RESULT 4
AW620463/C

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	FEATURES
AME20463	442 bp mRNA linear EST 03-DEC-2001	g1505603.y1	Gm-cl032	Glycine max CDNA clone	GENOME SYSTEMS CLONE ID:		
AME20463	Gm-cl032-1013 5', mRNA sequence.	AME20463					
AME20463.1	GI:7326665						
EST.							
soybean.							
Glycine max							
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.							
1 (bases 1 to 442)							
Shoemaker, R., Keim, P., Vodkin, L., Espelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, T., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Mattern, R. and Wilson, R.							
Public Soybean EST Project							
Unpublished (1999)							
Contact: Shoemaker R/Public Soybean EST Project							
Public Soybean EST Project							
Washington University School of Medicine							
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA							
Tel: 314 286 1800							
Fax: 314 286 1810							
Email: est@wustl.wustl.edu							
This clone is available through: Resgen, Invitrogen Corp., 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: c@resgen.com							
Insert Length: 1099	Std Error: 0.00						
High quality sequence stop: 436.							
Location/Qualifiers							

```

FEATURES
SOURCE

Location/Qualifiers
1..442
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1032-1013"
/clone_id="Gm-c1032"
/tissue_type="Cotyledons of 8-day-old 'Williams' seedlings"
/lab_host="DH10B"
/notice="Vector: Bluescript II XR; Site.1: EcoRI; Site.2: XhoI; This cDNA library was constructed from mRNA isolated from cotyledons of 8-day-old 'Williams' seedlings which were propagated on paper towels with distilled water for 3 days (etiolated), then greenhouse grown for 5 days in potting soil. The cotyledons were flash-frozen in liquid nitrogen. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGGAGAGAGAGAGACTGACG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-free, DNase-free water. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI from Promega (40U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then precipitated, redissolved, and ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). 100% of the white and 87.5% of the blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=28 and 8

```

BASE COUNT 78 a 180 c 90 g 94 t
ORIGIN

Query Match	37.68;	Score 38;	DB 10;	Length 442;
Best Local Similarity	67.98;	Pred. No. 32;		
Matches	53;	Conservative	0;	Mismatches 25;
			Indels	0;
			Gaps	0

QY 20 GGTCAATGGGCGGAGCAGGTGTGATCCGGCGTGGACCAAGCCGTGCAGGCCGATCGC 79
||||| ||| |||| | |||| |||| | |||| | |||| |
Db 292 GGTCAGAGCGCGGAGACAAGAGGTGTGAGCGCGCGGAGGAGGAGCGGCGGAGCGCGCTGC 233

QY 80 GCGGGGTTCCGGCGCG 97
 ||||| | |||||
 Db 232 GCGGAGGAGGAGTGC 215

RESULT 5
A0070009
20070000 368 bp mpDNA linear ESM 02-ADB-2002

DEFINITION	AB070009 Rice panicle shorter than 3cm Oryza sativa (japonica cultivar-group) cDNA clone E31777_9A, mRNA sequence.
ACCESSION	AB070009
VERSION	AB070009.1
GI	5004728

KEYWORDS	EST.
SOURCE	Oryza sativa (japonica)
ORGANISM	Oryza sativa (japonica)
	Eukaryota: Viridiplantae

REFERENCE
AUTHORS
Sasaki, T., and Yamamoto, K.
1 (bases 1 to 368)
Ehrhartoidae; Oryzae; Oryza.
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Oryzae; Oryza.

FILE RICE CDNA FROM PANICUM
JOURNAL Unpublished (1997)
COMMENT Contact: Takuji Sasaki
National Institute of

305-86602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468

```

FEATURES
  source      Location/Qualifiers
1. .368
PROJECT = 'RCP'

```

```

/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E31777_9A"

```

	BASE COUNT	55 a	116 c	145 g	49 t	3 others
/dev_stage="shorter than 3cm"						
/note="Organ: pancicle"						

Query Match	37.08;	Score 37.4;	DB 9;	Length 368;
Best Local Similarity	62.18;	Pred. NO. 44;		
...

[illegible]

QY 61 AAGCGTGACGGCATCGCGCGGGGTTCCGGCG 95
| | | | | | | | | | | | | | | | | | | |
Dh 206 ATGGCGCGCGCGCGCGTGGCGCGCGCGCGCTAGCGCG 24

RESULT 6	BO472223/c	LOCUS	DEFINITION
BO472223	634 bp	linear	EST 30-MAY-2007
HV04M06	HV	Hordeum vulgare CDNA clone HV04M06 5'-PRIME,	MRNA sequence.

[illegible]

AUTHORS	Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T., Chung,F. and Fraser,C.M.		
TITLE	ESTs from mycelia of Colletotrichum trifolii race 1		
JOURNAL	Unpublished (2002)		
COMMENT	Other-ESTs: EST532612		
	Contact: Deborah A. Samac		
	Department of Plant Pathology		
	University of Minnesota		
	495 Borling Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA		
	Tel: 612 625 1243		
	Fax: 651 649 5058		
	Email: debbye@puccini.crl.umn.edu		
	TIGR sequence name: MTSA159TK More information is available at:		
	www.medicago.org		
FEATURES	Seq primer: SKmod (CTA gaa cta gtc gat cc).		
SOURCE	Location/Qualifiers		
	1..657		
	/organism="Colletotrichum trifolii"		
	/strain="race 1"		
	/db_xref="taxon:5466"		
	/clone="pDSCR7-59"		
	/clone_1fb="DSCR"		
	/tissue_type="mycelia"		
	/dev_strage="Young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium containing 2%glucose)."		
	/lab_host="DH5alpha		
	/note="Vector: plusescript SK+; Site_1: EcoRI; Site_2: EcoRI; Isolate: 2sp2; cDNA was prepared from polyA+ enriched RNA The cDNA was ligated into lambda gtl from strichagene and packaged using Gigapack packaging extracts. An aliquot of the amplified library was used to transduce E. coli y1090 and phage DNA was purified from a liquid lysate. The cDNA inserts were gel purified after EcoRI digestion and ligated into plusescript SK+. Aliquots of the ligation were used to transform E. coli DH5alpha which were plated onto medium with X-gal for selection of recombinants."		
BASE COUNT	162 a	215 c	154 g
ORIGIN			126 t
Query Match	37.0%; Score 37.4; DB 14; Length 657;		
Best Local Similarity	62.1%; Pred. No. 43;		
Matches 59; Conservative	0; Mismatches 36; Indels 0; Gaps 0;		
Oy	1 CGTGGTCGATCGCGGCGGCATGAGCCGAGACAGAGTGTATCGCGGCGGAGC 60		
Db	593 CGGTGCTGAGAGTGGCGGTGTCGTATGAGCGGCTGGGGTGTGTGTCGGCGGCGGTG 534		
Oy	61 AAGCGTCCAGCGCATCGCGGCGGGGTTCGCGCGC 95		
Db	533 TGTGTGTGTGTCGCGGTCAAGCGGCGGCGGCGCGC 499		
RESULT 8			
LOCUS	BM500473	309 bp	mRNA linear EST 14-FEB-2002
DEFINITION	PAC0000000000531 Pioneer Af-1 array Zea mays cDNA, mRNA sequence.		
ACCESSION	BM500473		
VERSION	BM500473.1	GI:18659692	
KEYWORDS	EST.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC		
	clade; Panicoideae; Andropogoneae; Zea.		
REFERENCE	1 (bases 1 to 309)		
AUTHORS	Hunter,B.C., Beatty,M., Singletary,G., Hamaker,B., Larkins,B.A. and Jung,R.		
TITLE	Matze opaque endosperm mutations create extensive changes in patterns of gene expression		
JOURNAL	Unpublished (2002)		
COMMENT	Contact: Jung R		

FEATURES	source	location/Qualifiers
		1..642
		/organism="Triticum aestivum"
		/cultivar="Tatcher Lrl"
		/db_xref="taxon:4565"
		/clone="Patrl165C02"
		/clone_lib="Talrl1"
		/tissue_type="leaf tissue"
		/dev_stage="14 Days old"
		/lab_host="E. coli XL0LR"
		/vector="Vector: Lambda ZapII; mass excised in plasmid
		vector BKK-CMV (Stratagene).; Site.1: EcoRI; Site.2: XhoI
		mRNA obtained from wheat NIL "Tatcher Lrl 24 hours after
		inoculation with leaf rust pathogen Puccinia triticina
		race BBB carrying the avirulence gene Avr1."
BASE COUNT	138 a	212 c 209 g 82 t 1 others
ORIGIN		

Query Match	36.6%	Score 37	DB 12	length 696
Best Local Similarity	64.7%	Pred. No. 52		
Matches	55	Conservative	0	Mismatches 30
				Indels 0
				Gaps 0

QY	17	GCGGTCATATGCGCGGAGCAGTGTGTATCGCGGCGGAGCAAGCGGTACAGGCGAT	76
Db	135	GGAAGTGAAGAAGGAGGAGGAGCCCGCGCGACCTGCCCCCGCTGCAGAGAGAGCAGGCGAC	194
QY	77	GCGGCGGGGGTTCGCGGCGCGCGC	101
Db	195	GAGCGCGGCGGCGCGGCGGCGAAGCGC	219

[illegible]

TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex France Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES	Source 1..904 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CSODI061YB14" /clone_id="LPI_NFI006_PL2" /tissue_type="Placenta" /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI oligo(2) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT	138 a 87 c 444 g 179 t 56 others
ORIGIN	
Query Match	36.6%; Score 37; DB 9; Length 904;
Best Local Similarity	53.6%; Pred. Mismatches 52;
Matches	52; Conservative 10; N mismatches 35; Indels 0; Gaps 0;
Oy	4 GGGTCTCGATCGCGGCGGTCAATGGCCGCCAGACAGTGTGATCGCGGCGTGAGCAAG 63 Db 156 GKGTCGTCTGGTGGCGSCASATTGCAGCGCGCKGSGGGGSGGGGSGGSGG 215 Oy 64 CGGTGACGGCGCATCGCGCGGGTCTGGCGGCGCGCG 100 Db 216 GGTGTGGGGGGGGGGGGGGGGGGGGGGGGGGCG 252
RESULT 13	
B0669726/c	1065 bp mRNA linear EST 15-JUL-2002
LOCUS	
DEFINITION	AGENCOURT_8208383 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6254498
ACCESSION	B0669726
VERSION	B0669726
KEYWORDS	B0669726.1 GI:21780560
SOURCE	EST.
ORGANISM	human. Homo sapiens
REFERENCE	Eutheria; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1065) NIH-MGC http://mgc.ncl.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-jemail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (NLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/NLNL at: http://image.lnl.gov Plate: L1CM2404 row: b column: 03 Location/Qualifiers
AUTHORS	1..1065
TITLE	/organism="Homo sapiens"
JOURNAL	/db_xref="taxon:9606"
COMMENT	/clone="IMAGE:6254498"
	/clone_lib="NIH_MGC_102"
	/tissue_type="epidermoid carcinoma, cell line"

```

/lab_host="DH10B (phage-resistant)"
/site_1="XhoI;
/site_2="EcoRI; cdna made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAC(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH-MCC Library."
BASE COUNT      109 a      551 c      254 g      151 t
ORIGIN

```

Query Match	36.6%	Score 37:	DB 14;	Length 1065;
Best Local Similarity	67.5%	Pred. No. 51;		
Matches 52;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;

QY 24 AATGGGCGCGGAGCAGAGTGATGCGGCGGTGAGCAACGGGTGACGGGATTCGGCGG 83
 Db 629 AACGGGCGCGGGAGAGGGGGGAGCAGAGGGGTGGCGGTAAAGGGGGGGCGCGGGCGCGG 570
 QY 84 GGGTTTCGGCGCGCGCGG 100
 Db 569 GGGTCCGGGGATCGGG 553

RESULT 14	
AV182926/c	240 bp mRNA linear EST 21-JUL-1999
LOCUS	
DEFINITION	embryo Caenorhabditis elegans cDNA, clone YK644d3 3', mRNA sequence.
ACCESSION	AV182926
VERSION	AV182926.1 GI:5562827
KEYWORDS	EST.
	Caenorhabditis elegans.

REFERENCE

AUTHORS

Eunaryota, Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea
; Rhabditidae; Pelodermidae; Caenorhabditis.
1 (bases 1 to 240)

Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H.,
Nishizaki, A., Mochizuki, T., Zeng, Q., Watanabe, H., Sugimoto, A., Sano,
M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and
NOMOTO, H.

TITLE Expressed genes in *C.elegans*
JOURNAL Unpublished (1999)
COMMENT Contact: Yuji Kohara
 Genome Biology Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

FEATURES	Location/Qualifiers
SOURCE	1. .240

```

/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk644d3"
/clone_1lb="Yui Kohara unpublished CDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
44 a 115 c 36 g 38 t 7 others
BASE COUNT
ORIGIN

```

Query Match	36.4%	Score 36.8	DB 9	Length 240
Best Local Similarity	63.6%	Pred. No. 61		
Matches 56	Conservative 0	Mismatches 32	Indels 0	Gaps 0

Oy 5 GGTTCGATCGCGCGGTCAAATGGGCCGAGCAGTGTGATCCGGCGGTGGAGCAAGC 64
 ||| || ||||| || || ||||| || || ||||| ||
 Db 223 GGTTCGAGCGCGCGGTGAGAGTGTGTGGCGGCGGTGCGTATGGAAGCGGTGGAGAGATGG 164

QY 65 GGTGCAGCGCATCGCGGCGGGGTTCCGG 92

Db 163 GGTGGCGCGGTGGCGGACAGATCTG 136

RESULT 15	
AV181002/c	
LOCUS	300 bp
DEFINITION	mRNA linear EST 21-JUL-1999
	AV181002 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
	embryo Caenorhabditis elegans cDNA clone yk61g12 3', mRNA
	sequence.

ACCESSION	AV181002
VERSION	AV181002.1
KEYWORDS	GI:5560903
SOURCE	EST.
ORGANISM	<i>Caenorhabditis elegans</i>

REFERENCE
AUTHORS
1 (bases 1 to 300)
Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H., Nishiyaki, A., Motolashvili, T., Zeng, Q., Matanabe, H., Sugimoto, A., Sano, M., Miyata, A., Mitani, Y., Ida, K., Uesugi, H., Sugiyama, Y. and Nomoto, H.
TITLE
JOURNAL
COMMENT
Expressed genes in C.elegans
Unpublished (1999)
Contact: Yuji Kohara

Genome Biology Lab.
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel.: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykoba@nig.nig.ac.jp.
location/Qualifiers
1.300

BASE COUNT	ORIGIN
72 a	120 c 48 g 52 t 8 others

Query Match	36.4%	Score 36.8	DB 9	Length 300
Best Local Similarity	63.6%	Pred. No. 60		
Matches 56; Conservative	0	Mismatches 32	Indels 0	Gaps 0

QY	65	GGTGACGCGCATCCGCGCGGGGTTCCGCG	92
Db	218	GGTGCGCGCGGTGGGGGCGACGATCTGtg	191

Search completed: November 5, 2002, 16:17:58
Job time : 1118.44 secs

Search completed: November 5, 2002, 16:17:58
Job time : 1118.44 secs


```
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DEIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US 09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,703
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 3936
LENGTH: 401
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007249.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.7
US-09-864-761-3936
Query Match 34.7%; Score 35; DB 10; Length 401;
Best Local Similarity 61.5%; Pred. No. 0.88;
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 5 GGTGTGATGCGGCGGTCATGGCGGAGCAGTGTGATGCGGCGGCTGAGCAAGC 64
DB 142 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 201
QY 65 GTGACAGCGCATGCGCGGGGTTCGCGGCG 95
```

```
DB 202 GATGTGCGCGGTGCGAGTGTGCGGTGTGTGCGG 232
RESULT 3
US-09-864-761-20699
Sequence 20699, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DEIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US 09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20699
LENGTH: 446
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007249.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.7
OTHER INFORMATION: NT HIT: AL161539.2, EVALUATE 3.70e+00
```



```
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 19241
;; LENGTH: 1075
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL078472.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
;; OTHER INFORMATION: EST_HUMAN HIT: A7739739.1, EVALUATE 1.00e+00
;; OTHER INFORMATION: NT HIT: AL163201.2, EVALUATE 2.00e-19
US-09-864-761-19241

Query Match      32.7%; Score 33; DB 10; Length 1075;
Best Local Similarity 60.7%; Pred. No. 2.5;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 5 GGTGTCGATCGCGGCGTCAATGCGCGGAGCAAGGTGATGCGGCGGTGAGCAAGC 64
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 103 GGTAGTGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 162
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 65 GGTGACAGCGATCGCGGCGGCGGCTTCGCGG 93
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 191
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-09-864-761-2513
; Sequence 2513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Penn, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
```

```
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 2513
;; LENGTH: 1403
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL078472.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
US-09-864-761-2513

Query Match      32.7%; Score 33; DB 10; Length 1403;
Best Local Similarity 60.7%; Pred. No. 2.5;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 5 GGTGTCGATCGCGGCGTCAATGCGCGGAGCAAGGTGATGCGGCGGTGAGCAAGC 64
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 453 GGTAGTGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 512
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 65 GGTGACAGCGATCGCGGCGGCGGCTTCGCGG 93
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 513 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 541
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-880-107-2300
; Sequence 2300, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darcia T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
```

TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 FILE REFERENCE: 44921-5028-WO
 CURRENT APPLICATION NUMBER: US/09/880,107
 PRIOR FILING DATE: 2001-06-14
 PRIOR APPLICATION NUMBER: US 60/211,379
 PRIOR FILING DATE: 2000-06-14
 PRIOR APPLICATION NUMBER: US 60/237,054
 PRIOR FILING DATE: 2000-10-02
 NUMBER OF SEQ ID NOS: 3950
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO 2300
 LENGTH: 3715
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: Genbank Accession No. US20020142981A1 M23263
 US-09-880-107-2300

Query Match	32.5%	Score 32.8	DB 10	Length 3715
Best Local Similarity	64.5%	Pred. No. 2.7		
Matches 49	Conservative 0	Mismatches 27	Indels 0	Gaps 0

DQ 20 GGTCAATGGGCGCAGACAGGTGGTGCATCCGGGGCGCTGGAGCAAGCGGTGCACGCCGATCC 79
|| || | ||| | ||| | ||| | ||| | ||| |
Db 1843 GGCCAGTTGTATGACCCGTCGTGCTGTCGGGGGTGGTGGTGCGCGCGCGGCGACCGCC 1902

QY	80	GCGGGGTTCCGGCG	95
Db	1903	GCGGGGCGCGCGCG	1918

RESULT 10
US-09-893

; Sequence 327, Application US/09893737
; Patent No. US2002011085A1

```

; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.

```

```

; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41

```

; CURRENT APPLICATION NUMBER: US/09/893,737
 ; CURRENT FILING DATE: 2001-06-28

;; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30

```
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0.0
```

```

; SEQ ID NO 327
; LENGTH: 681

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
; FEATURE:
; NAME/KEY: CDS
```

LOCATION: (1)...(681)
US-09-893-737-327

Query Match	32.3%	Score 32.6	DB 10	Length 681
Best Local Similarity	63.3%	Pred. 30.3	2	
Matches 50	Conservative	0	Mismatches 29	Indels 0
				Gaps 0

0y 16 CCGGGTCATGGGCGGAGCAGGTGTGATCGGGCGTGAGCAAGCGGTGCAGCGGA 75
| | | | | | | | | | | | | | | |
Db 461 CTGAGTAGCCGAGCAGCAGCAGCGGGCCGGGGCAGCCTTGAGCAGGCGGCAAGCAGGGC 402

QY 76 TCGCGCGGGGTTCCGGC 94
 11 111111 11 111
 Db 401 GCGGCGCGGGGTCGGGC 381

RESULT 11

US-09-969-708-328
; Sequence 328, Application US/09969708

; Patent No. US20020102532A1
; GENERAL INFORMATION:

```

? APPLICANT: Augustus, Meena
? TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sigm
? TITLE OF INVENTION: Sets
? FILE REFERENCE: 689290-70
? CURRENT APPLICATION NUMBER: US/09/969,708
? CURRENT FILING DATE: 2001-10-03
? PRIOR APPLICATION NUMBER: US/60/237,606
? PRIOR FILING DATE: 2000-10-03
? PRIOR APPLICATION NUMBER: US/60/237,608
? PRIOR FILING DATE: 2000-10-03
? PRIOR APPLICATION NUMBER: US/60/237,425
? PRIOR FILING DATE: 2000-10-03
? NUMBER OF SEQ ID NOS: 658
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 328
? LENGTH: 1797
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-969-708-328

```

Query Match	31.9%	Score	32.2	DB	10	Length	1797
Best Local	Similarly	Pred.	No. 3.8				
Matches	55	Conservative	0	Mismatches	38	Indels	0
						Gaps	0

QY 4 GGGTGTGCAATCGCGGGCTCAATGGGCGGAGCAGGTGTGATCGCGGGCGTGGAGCAAG 63
 |||||
 Db 1089 GCGCTGGCGGCCACCAAGTGGCGGAAGCGGAAGCTGAGCGGCAATCGCGGCCCTGGAGAGACA 1148

Oy	64	CGGTGACGGCGATCCGGCGGGGGTTCGCGGCGC	96
Db	1149	AGGTGAGACCGCTCAAGCCGAGAACGCGGGGC	1181

RESULT 12
US-09-815

; Sequence 7796, Application US/09815242
; Patent No. US20020061569A1

; GENERAL INFORMATION;
; APPLICANT: Haselbec

```

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.

```

```

; APPLICANT: wall, daniel
; APPLICANT: Trawick, John D.

```

```

; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

```

```

; APPLICANT:  XU, H. Howard
; TITLE OF INVENTION:  Identification of Essential Genes in

```

```
; TITLE OF INVENTION: Prokaryotes
;
; FILE REFERENCE: ELITRA.011A
```

```

; CURRENT APPLICATION NUMBER: US/0
; CURRENT FILING DATE: 2001-03-21

```

;; PRIOR APPLICATION NUMBER: 60/191,07
; PRIOR FILING DATE: 2000-03-21

;
; PRIOR APPLICATION NUMBER: 60/206,84
;
; PRIOR FILING DATE: 2000-05-23
;

;
; PRIOR APPLICATION NUMBER: 60/207,72
; PRIOR FILING DATE: 2000-05-26

;
; PRIOR APPLICATION NUMBER: 60/242,57
; PRIOR FILING DATE: 2000-10-23
;

;
; PRIOR APPLICATION NUMBER: 60/253,62
; PRIOR FILING DATE: 2000-11-27
;

;
; PRIOR APPLICATION NUMBER: 60/257,93
; PRIOR FILING DATE: 2000-12-22

;
; PRIOR APPLICATION NUMBER: 60/269,30
; PRIOR FILING DATE: 2001-02-16
;

```
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 7796
; LENGTH: 939

```

```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

```

```
; FEATURE:
; NAME/KEY: CDS
```

LOCATION: (1)...(939)
US-09-815-242-7796

Query Match 31.7%; Score 32; DB 10; Length 939;
Best Local Similarity 62.5%; Pred. No. 4.4;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 22 TCATGCGCGGACAGTGTGATCGGGCGGTGAGCAAGCGGTGCGATCGCGG 81
DB 485 TCACGCGCGCGGCGCGGCGGTGATCGCGGTGCGCGGTGAGCGGTGCGATCG 544
OY 82 CGGGGTTCGCGCGCGCGCGG 101
DB 545 AGGATGCAAGCGCGCGCGG 564

RESULT 13

US-09-735-705-215/C
Sequence 215, Application US/09735705
Patent No. US20020052329A1
GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aljun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
APPLICANT: Fanger, Neil
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C14
CURRENT APPLICATION NUMBER: US/09/735.705
CURRENT FILING DATE: 2000-12-12
NUMBER OF SEQ ID NOS: 419
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 215
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (17)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (20)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (60)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (61)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (365)
OTHER INFORMATION: n-A,T,C or G
US-09-735-705-215

Query Match 31.5%; Score 31.8; DB 10; Length 381;
Best Local Similarity 61.4%; Pred. No. 5.1;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 19 CGGTCATGCGCGGACAGTGTGATCGGGCGGTGAGCAAGCGGTGCGATCG 78
DB 231 CGTCCCTGCGCGCGGTGATCGCGGTGATCGCGGTGATCGCGGTGATCG 172
OY 79 CGGCGGGGTTCGCGCGCGCGCGG 101
DB 171 CGGCGGGGGGGCGCGCGCGCGG 149

RESULT 14

US-09-850-716A-215/C
Sequence 215, Application US/09850716A
Patent No. US20020115139A1
GENERAL INFORMATION:
APPLICANT: Kalos, Michael D.
APPLICANT: McNeill, Patricia D.
APPLICANT: Retter, Marc W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C15
CURRENT APPLICATION NUMBER: US/09/850.716A
CURRENT FILING DATE: 2001-05-07
NUMBER OF SEQ ID NOS: 440
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 215
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (17)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (20)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (60)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (61)
OTHER INFORMATION: n-A,T,C or G
NAME/KEY: unsure
LOCATION: (365)
OTHER INFORMATION: n-A,T,C or G
US-09-850-716A-215

Query Match 31.5%; Score 31.8; DB 10; Length 381;
Best Local Similarity 61.4%; Pred. No. 5.1;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 19 CGGTCATGCGCGGACAGTGTGATCGGGCGGTGAGCAAGCGGTGCGATCG 78
DB 231 CGTCCCTGCGCGCGGTGATCGCGGTGATCGCGGTGATCGCGGTGATCG 172
OY 79 CGGCGGGGTTCGCGCGCGCGCGG 101
DB 171 CGGCGGGGGGGCGCGCGCGCGG 149

RESULT 15

US-09-897-778-215/C
Sequence 215, Application US/09897778
Patent No. US20020147143A1
GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Fanger, Gary R.
APPLICANT: Vedrick, Thomas S.
APPLICANT: Carter, Darick
APPLICANT: Watanabe, Yoshihiro
APPLICANT: Henderson, Robert A.
APPLICANT: Peckham, David W.
APPLICANT: Fanger, Neil
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C16
CURRENT APPLICATION NUMBER: US/09/897.778
CURRENT FILING DATE: 2001-06-28
NUMBER OF SEQ ID NOS: 467
SOFTWARE: FastSeq for Windows Version 4.0

```

? SEQ ID NO 215
? LENGTH: 381
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURES
? NAME/KEY: misc_feature
? LOCATION: 17, 20, 60, 61, 365
? OTHER INFORMATION: n = A,T,C or G
? OS=98-897 -778-215

```

Query Match	31.5%	Score 31.8	DB 10	Length 381
Best Local Similarity	61.4%	Pred. No. 5.1		
Matches 51, Conservative	0	Mismatches 32	Indels 0	Gaps 0

[illegible]

Search completed: November 5, 2002, 23:00:14
Job time : 89.2537 secs

Patent No. 6121029
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/335,409
: CURRENT FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 60
|||
Db 23883 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 23942
QY 61 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 101
|||
Db 23943 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 23983

RESULT 3
US-09-568-102-1
: Sequence 1, Application US/09568102
: Patent No. 6346404
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,102
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 60
|||
Db 23883 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 23942
QY 61 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 101
|||
Db 23943 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 23983

RESULT 4
US-09-567-969-1

Sequence 1, Application US/09567969
: Patent No. 635457
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,969
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 60
|||
Db 23883 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 23942
QY 61 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 101
|||
Db 23943 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 23983

RESULT 5
US-09-568-480-1
: Sequence 1, Application US/09568480
: Patent No. 635458
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,480
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 60
|||
Db 23883 CGTGGGTTCATCCGCGCGGTCAATGGGCGGACAGGTGTGATCCGGCGGTGGAGC 23942
QY 61 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 101
|||
Db 23943 AAGCGGTGACAGCGATCCGCGCGGCGGTTTCGCGCGCGCGC 23983


```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6353459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGTGGGTGTCGATCCGCGCGGTGATGGCCGAGCAGGTGATCGCGGCGTGGAGC 60
Db 23883 CGTCGTCGTCATCCGCGCGGTGATGGCCGAGCAGGTGATCGCGGCGTGGAGC 23942
OY 61 AAGCGGTGACAGCGATCGCGGCGGTTCCGCGCGCGCGC 101
Db 23943 AAGCGGTGACAGCGATCGCGGCGGTTCCGCGCGCGCGC 23983

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGTGGGTGTCGATCCGCGCGGTGATGGCCGAGCAGGTGATCGCGGCGTGGAGC 60
Db 23883 CGTCGTCGTCATCCGCGCGGTGATGGCCGAGCAGGTGATCGCGGCGTGGAGC 23942
OY 61 AAGCGGTGACAGCGATCGCGGCGGTTCCGCGCGCGCGC 101
Db 23943 AAGCGGTGACAGCGATCGCGGCGGTTCCGCGCGCGCGC 101
```

```
Db 23943 AAGCGGTGACAGCGATCGCGGCGGTTCCGCGCGCGCGC 23983

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 6.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CGTGGGTGTCGATCCGCGCGGTGATGGCCGAGCAGGTGATCGCGGCGTGGAGC 60
Db 23883 CGTCGTCGTCATCCGCGCGGTGATGGCCGAGCAGGTGATCGCGGCGTGGAGC 23942
OY 61 AAGCGGTGACAGCGATCGCGGCGGTTCCGCGCGCGCGC 101
Db 23943 AAGCGGTGACAGCGATCGCGGCGGTTCCGCGCGCGCGC 23983

RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,4380S1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSBQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match          45.9%; Score 46.4; DB 4; Length 15872;
Best Local Similarity 67.7%; Pred. No. 0.0058;
Matches 65; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 6 GTGTGATCGCGCGGTGATGGCCGAGCAGGTGATCGCGGCGTGGAGCAGCG 65
Db 10574 GTGTGATCGCGCGGTGATGGCCGAGCAGCGGTGATCGCGGCGTGGAGCAGCG 10633
OY 66 GTGACGAGCGATCGCGGCGGTTCCGCGCGCGCGC 101
Db 23943 GTGACGAGCGATCGCGGCGGTTCCGCGCGCGCGC 101
```

Db 10634 GTGCCGAGGCGGCGGCAGCTCGCCGACGAGGC 10669

```

RESULT 10
US-08-804-227C-1
Sequence 1, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 15/11
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1
Query Match 45.9%; Score 16.4; DB 2; Length 43280;
Best Local Similarity 67.7%; Pred. No. 0.0057;
Matches 65; Conservative 0; Mismatches 31; Indels 0; Gaps 0.
Qy 6 GTGTCGATCGCGGGGTCATATGCGCGGAGCAGTGTGATGATCGCGGCGTGGAGCAAGCG 65
Db 38436 GTGTCGCGTCGCGCGGCGGTGAACGGCCCGCGTCGTGTGTGATGATTCGCGGCGGAGGAAGCG 38435
Oy 66 GTGCAAGCGGATCGCGGCGGCGGTTGCGGCGCGCGCGC 101
Db 38496 GTGGCCGAGGCGCGCGCGCACACTCCGCGGAGCGAGCG 38531

```

```

RESULT 11
US-09-103-840A-2
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match          42.8%; Score 43.2; DB 4; Length 4403765;
Best Local Similarity 65.6%; Pred. No. 0.028;
Matches 63; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY      6 GTGTGATCGCGCGCGTCAATGGCCGAGCAGTGATGATCGCGCGCGTGGAGCAAGCG 65
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 489502 GTGACATCGCGCGCGTCAATGGTCCGGCTTGATGATGATCGCGCGCGTGGAGCAAGCG 489561

QY      66 GTGACAGCGATCGCGCGCGGCGTTCGCGCGCGCGCGCG 101
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 489562 GTGACGCGCATCGCTGATCGCTGCGCGCGCGCGCGCG 489597

RESULT 12
US-09-103-840A-1
/ Sequence 1, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          42.8%; Score 43.2; DB 4; Length 4411529;
Best Local Similarity 65.6%; Pred. No. 0.028;
Matches 63; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY      6 GTGTGATCGCGCGCGTCAATGGCCGAGCAGTGATGATCGCGCGCGTGGAGCAAGCG 65
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 488060 GTGACATCGCGCGCGTCAATGTCGCGCTTGATGATGATCGCGCGCGTGGAGCAAGCG 488119

QY      66 GTGACAGCGATCGCGCGCGGCGTTCGCGCGCGCGCGCG 101
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 488120 GTGACGCGCATCGCTGATCGCTGCGCGCGCGCGCGCG 488155

```

```
RESULT 13
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kustoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
; US-08-804-227C-13

Query Match          41.0%; Score 41.4; DB 2; Length 13987;
Best Local Similarity 63.6%; Pred. No. 0.079; 36; Indels 0; Gaps 0;
Matches 63; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 GTGGGTGTCATCGCGCGGTCATATGGCGCGAGCAGTGTGTATCGCGGGCGGTGAGCA 61
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 5635 GTGGGTGTCGTGGCGCGGTCATATGGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGG 5694
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 62 AGCGGTGACAGGATCGCGCGGGGTTCGCGCGCGCG 100
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 5695 TGTGCTGAGATCGGTGTCTCGCTGATGGGGGATGG 5733
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 14
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kustoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
```

```
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match          41.0%; Score 41.4; DB 2; Length 44377;
Best Local Similarity 63.6%; Pred. No. 0.077; 36; Indels 0; Gaps 0;
Matches 63; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 GTGGGTGTCATCGCGCGGTCATATGGCGCGAGCAGTGTGTATCGCGGGCGGTGAGCA 61
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2437 GTGGGTGTCGTGGCGCGGTCATATGGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGG 2496
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 62 AGCGGTGACAGGATCGCGCGGGGTTCGCGCGCGCG 100
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 2497 TGTGCTGAGATCGGTGTCTCGCTGATGGGGGATGG 2535
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 15
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kustoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
```

STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,138
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match 41.08; Score 41.4; DB 2; Length 44377;
Best Local Similarity 63.68; Pred. No. 0.077;
Matches 63; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 2 GTGGGTCTCATCGCGCGGTCAATGGGCCAGACAGGTGATCGCGCGGTGAGCA 61
|||||
Db 2437 GTGGGTCTCGGTTCGCGCGGTGATGGGCCAGGTGCTGTTGTCGGGTGATGTGGG 2496
|||||
QY 62 AGCGGTGCAGGCGATCGCGCGGGGTTCGCACCGCGCGG 100
|||||
Db 2497 TGTCTGGAGTCGGGTGCTGCTGATAGGGGATGG 2535
|||||

Search completed: November 5, 2002, 15:36:40
Job time: 1926.74 secs


```
FT      /*tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /*tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /*tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /*tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /*tag= h
FT      /label= epob_gene
FT      /note= "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /*tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /*tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /*tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /*tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /*tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /*tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /*tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /*tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /*tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /*tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /*tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /*tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /*tag= w
FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /*tag= x

FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /*tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /*tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /*tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /*tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /*tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /*tag= ad
FT      /label= epod_gene
FT      /note= "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /*tag= ae
FT      /note= "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /*tag= af
FT      /note= "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /*tag= ag
FT      /note= "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /*tag= ah
FT      /note= "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /*tag= ai
FT      /note= "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /*tag= aj
FT      /note= "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /*tag= ak
FT      /note= "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /*tag= al
FT      /note= "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /*tag= am
FT      /note= "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /*tag= an
FT      /note= "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /*tag= ao
FT      /note= "encodes DH5"
FT      misc_RNA
FT      29433..30287
FT      /*tag= ap
FT      /note= "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /*tag= aq
FT      /note= "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /*tag= ar
FT      /note= "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /*tag= as
FT      /note= "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /*tag= at
FT      /note= "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /*tag= au
```

```
FT FT /note= "encodes DH6"
FT FT 35760..36641
FT FT /*tag= av
FT FT /note= "encodes ER6"
FT FT 36705..37256
FT FT /*tag= aw
FT FT /note= "encodes KR6"
FT FT 37470..37769
FT FT /*tag= ax
FT FT /note= "encodes ACP6"
FT FT 37912..49308
FT FT /*tag= ay
FT FT /label= epoE_gene
FT FT /note= "encodes modules 7 and 8"
FT FT 38014..39375
FT FT /*tag= az
FT FT /note= "encodes KS7"
FT FT 39589..40626
FT FT /*tag= ba
FT FT /note= "encodes AT7"
FT FT 41341..41922
FT FT /*tag= bb
FT FT /note= "encodes KR7"
FT FT 42181..42423
FT FT misc_RNA

Query Match 100.0%; Score 101; DB 21; Length 71969;
Best Local Similarity 100.0%; Pred. No. 1,1e-14;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CGTGGGTGTCGATCCGCGCGGTCAATGGGCGGACAGGTGTCGATCCGCGCGGTGAGC 60
Db 18271 CGTGGGTGTCGATCCGCGCGGTCAATGGGCGGACAGGTGTCGATCCGCGCGGTGAGC 18330
Oy 61 AAGCGGTGACGCGCATCCGCGCGGTTCGCGCGCGCGCGGC 101
Db 18331 AAGCGGTGACGCGCATCCGCGCGGTTCGCGCGCGCGCGGC 18371

RESULT 2
AA25587
ID AA25587 standard; DNA; 68750 BP.
XX
AC AA25587;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT 1..1826
FT CDS /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAV58581)"
FT 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT 6226..6675
FT /*tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT 7610..11875
FT /*tag= f
```

```
FT FT /product= "Type I polyketide synthase, EPOS A
FT FT 11872..116104
FT FT /*tag= g
FT FT /product= "Non-ribosomal peptide synthetase, EPOS P
FT FT 16251..21749
FT FT /*tag= h
FT FT /product= "Type I polyketide synthase, EPOS B
FT FT 21746..43519
FT FT /*tag= i
FT FT /product= "Type I polyketide synthase, EPOS C
FT FT 43524..54920
FT FT /*tag= j
FT FT /product= "Type I polyketide synthase, EPOS D
FT FT 54935..62254
FT FT /*tag= k
FT FT /product= "Type I polyketide synthase, EPOS E
FT FT 62369..63628
FT FT /*tag= l
FT FT /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT FT 63779..64333
FT FT /*tag= m
FT FT /product= "Orf 6 protein (AAV58585)"
FT FT complement (63853..64290)
FT FT /*tag= n
FT FT /product= "Orf 7 protein (AAV58586)"
FT FT 64363..64920
FT FT /*tag= o
FT FT /product= "Orf 8 protein (AAV58587)"
FT FT complement (64287..64727)
FT FT /*tag= p
FT FT /product= "Orf 9 protein (AAV58588)"
FT FT 65063..65767
FT FT /*tag= q
FT FT /product= "Orf 10 protein"
FT FT complement (65008..65874)
FT FT /*tag= r
FT FT /product= "Orf 11 protein (AAV58590)"
FT FT complement (65871..66338)
FT FT /*tag= s
FT FT /product= "Orf 12 protein (AAV58591)"
FT FT 66667..67137
FT FT /*tag= t
FT FT /product= "Orf 13 protein (AAV58592)"
FT FT 67334..68251
FT FT /*tag= u
FT FT /product= "Orf 14 protein (AAV58593)"
FT FT 68346..68750
FT FT /partial
FT FT /*tag= v
FT FT /product= "Partial Orf 15 protein (AAV58594)"
FT FT /note= "No termination codon given in the specification"
FT FT WO996028-A2.
FT FT 23-DEC-1999.
FT FT 16-JUN-1999; 99WO-EP04171.
FT FT 18-JUN-1998; 98US-0099504.
FT FT 24-SEP-1998; 98US-0101631.
FT FT 05-FEB-1999; 99US-0118906.
FT FT (NOVS ) NOVARTIS AG.
FT FT (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
FT FT Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
```

XX	WPI: 2000-097741/08.
DR	P-PsDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR	AAY58592, AAY58593, AAY58594.
XX	
PT	New isolated epoethiolone synthase genes, used for the recombinant
PT	production of epoethiolone for use in cancer therapy
PS	Claim 14; Page 87-104; 174pp; English.
XX	
CC	This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC	comprising 22 open reading frames (ORFs) and includes genes encoding
CC	proteins involved in the biosynthesis of epoethiolones. Epoethiolones A and
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC	starter unit; polyketides being synthesised from two-carbon building
CC	blocks, the beta-carbon of which always carries a keto group. Each round
CC	of two-carbon addition is carried out by a complex of enzymes known as
CC	the polyketide synthase. In a manner similar to fatty acid biosynthesis,
CC	EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC	the thiazolidine ring formation of epoethiolones, and EPOS B, EPOS C, EPOS D
CC	and EPOS E (AAY58575-158578) are involved in polyketide backbone
CC	formation. EPO F (AAY58579) is an epoethiolone macro lactonase, and
CC	the proteins Orf 3 (AAY58582) and Jrf14 (AAY58593) are thought to be
CC	involved in transport. Epoethiolones mimic the biological activity of
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic
CC	compositions. Epoethiolones exhibit a much lower drop in potency against a
CC	multiply drug-resistant cell line compared with taxol, and are
CC	considerably less efficiently exported from such cells by the multidrug
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC	epoethiolones as anticancer agents, they are problematical to produce on a
CC	large scale. Sorangium cellulosum is difficult to ferment, producing
CC	poor yields of epoethiolones. The nucleic acids of the invention may be
CC	used for the recombinant production of epoethiolones in a heterologous host
CC	that is more amenable to fermentation.
SQ	Sequence 68750 BP: 9596 A; 22458 C; 2537 G; 11159 T; 0 other:
	Query Match 98.4%; Score 99.4; DB 21; Length 68750;
	Best Local Similarity 99.0%; Pred. No. 2, 6e-14;
	Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0
OY	1 CGTGGGTGCGATCGCGGGGCTCAAGGC CGAGCAGTGTCGATCGCGCGTGAGC 60
DB	23883 GTTCGGTGTGATCGCGGGGGTCAATGGG CGAGCAGTGTCGCGCGTGAGC 23942
OY	61 AAGCGGTGCGATCGCGGGGCTTCGCGCGCGCGCG 101
DB	23943 AAGCGGTGCGATCGCGGGGCTTCGCGCGCGCG 23983
RESULT 3	
ID	AAF24892 standard; DNA; 20394 BP.
XX	AAF24892;
AC	
DT	20-APR-2001 (first entry)
XX	
DE	Pimaricin biosynthesis associated polyketide synthase gene.
XX	
KW	Polyketide synthase; oxidative modification; location; metabolite; antibiotic;
XX	anticancer; pimaricin; ss.
OS	Streptomyces natalensis.
FT	Key Location/Qualifiers
FT	CDS 1..20394
FT	/tag= a
FT	/product= "polyketide synthase"

PN		WO200077222-A1.	
PP		21-DEC-2000.	
PQ		14-JUN-2000; 2000MO-EP06227.	
PR		14-JUN-1999; 99EP-0201893.	
PS		(STAM) DSM NY.	
PT		Martin JF, Aparicio JF, Colina AJ;	
PV		WPI: 2001-080693/09.	
PX		P-PsDB; AAB31358.	
PY		New polynucleotides encoding enzymes involved in the biosynthesis of pimarinin, useful for modifying the biosynthesis of pimarinin and in the synthesis of new compounds -	
PZ		Disclosure: page 53-80; 116pp; English.	
QA		The present sequence encodes a polyketide synthase which is associated with the biosynthesis of pimarinin. The polyketide synthase polypeptide is useful for the oxidative modification of a methyl group of a suitable compound, e.g. a bioactive compound including a secondary metabolite, antibiotics and anticancer agents. Recombinant cells comprising the gene are useful for the production of pimarinin. The polyketide synthase polynucleotide may be over expressed in Streptomyces, leading to an increase in the biosynthesis of pimarinin, as a source of primers for amplification reaction and as probes.	
QB		Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;	
QC		Query Match 59.6%; Score 60.2; DB 22; Length 20394;	
QD		Best Local Similarity 76.3%; Pred. NC. 2.2e-05;	
QE		Matches 74; Conservative 0; Mismatches 23; Indels 0; Gaps 0;	
QF	4	GCGTGTGATCCGGCGCTCATTTGGGCCGAGCAGTGATTCGGCGCTGGACCAAG 63	
QG	Db	11918 GCGTCTCGCTCCGGCGGCTCATTTGGGCCGCTGTGATTCGCGGCGGAGAGC 11977	
QH	Qy	64 CGGTGCAGCGCATTCGCGCGGGGTTTCGCGCGCGCGG 100	
QI	Db	11978 CGGTCTGCGCATTCGCGCGGCGACTTCGCGGGGAGG 12014	
QJ		RESULT 4	
QK		AAD17185	
QL	ID	AAD17185 standard; DNA; 27541 BP.	
QM	AC	AAD17185;	
QN	XX	29-NOV-2001 (first entry)	
QO	XX	Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.	
QP	XX	Polyketide synthase; PKS; macrocyclic; nystatin; PKS gene cluster;	
QQ	KW	antifungal; antibiotic; nys2; ds.	
QR	XX	Streptomyces noursei.	
QS	OS		
QT	Key	Location/Qualifiers	
QU	CDS	complement (454..1191)	
QV	FT	/tag= a	
QW	FT	/product= "NysF protein"	
QX	FT	/note= "CDS does not include start codon"	
QY	FT	complement (1275..3092)	
QZ	FT	/tag= b	
RA	FT	/product= "NysG protein"	
RB	FT	complement (3070..4824)	
RC	FT	/tag= c	
RD	FT	/note= "CDS does not include start codon"	
RE	FT	/product= "NysH protein"	


```

FT CDS 5122..6156 /*tag= d
FT /product= "Nysd3 protein"
FT CDS 6338..27541 /*tag= e
FT /product= "NysI partial protein"
FT /note= "Cds does not include stop codon"
XX
PN W0200159126-A2.
XX
XX 16-AUG-2001.
XX PD
XX PE 08-FEB-2001; 2001WO-GB00509.
XX PR 08-FEB-2000; 2000GB-0002840.
XX PR 10-APR-2000; 2000GB-0008786.
XX PR 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNPE) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DEIE/) DIEGLEMSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FUAE/) FUJAEVIK E.
PA (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
PI Zotchev SP, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sielta H, Gulliksen O;
DR WP1: 2001-557614/62.
DR P-PSDB; AAEL0138, AAEL0139, AAEL0140, AAEL0141, AAEL0142.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
XX Claim 2; Page 151-166; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a streptomycetes noursei nys2 DNA of nystatin PKS gene cluster.
SO Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match 50.7%; Score 51.2; DB 22; Length 27541;
Best Local Similarity 70.8%; Pred.No. 0.0024;
Matches 68; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 5 GGTGTCGATCGGGGGGTCATGCGGCCGAGCAAGTGTCGATCGGGCGCTGGACGAAC 64
||| ||||||| ||| ||||||| ||| ||||||| ||| ||||||| ||| |||
13813 GTGTGCATCGCGCCGCTCAACGGGCCCCAGTCGCTGTGATCTCCGTGACGAGAGAGC 13872
OY 65 GGTGCAAGCATCGCGGGGTTGCGGCGCGCG 100
| | ||||||| ||| ||||||| |||
Db 13873 CGCCGAACGATCGCGCCACCCTTCCGACGCG 13908

RESULT 5
AADI17186 ID AADI17186 standard; DNA; 125401 BP.
AC AADI17186;
XX
XX 29-NOV-2001 (first entry)
XX DE Streptomycetes noursei nystatin PKS gene cluster DNA.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; ds.
XX
```

XX	OS	Streptomyces	noursei.	
XX	Key	Location/Qualifiers		
XX	Key	6337..34771		
XX	CDS	/tag= a		
XX	CDS	/product= "NysI complete protein"		
XX	CDS	34792..51099		
XX	CDS	/tag= b		
XX	CDS	/product= "NysJ protein"		
XX	CDS	51155..57355		
XX	CDS	/tag= c		
XX	CDS	/product= "NysK protein"		
XX	CDS	57303..58687		
XX	CDS	/tag= d		
XX	CDS	/product= "NysL protein"		
XX	CDS	complement (58786..58980)		
XX	CDS	/tag= e		
XX	CDS	/product= "NysM protein"		
XX	CDS	/note= "CDS does not include start codon"		
XX	CDS	complement (59045..60241)		
XX	CDS	/tag= f		
XX	CDS	/product= "NysN protein"		
XX	CDS	/note= "CDS does not include start codon"		
XX	CDS	complement (60238..61296)		
XX	CDS	/tag= g		
XX	CDS	/product= "NysD2 complete protein"		
XX	CDS	120628..121308		
XX	CDS	/tag= h		
XX	CDS	/product= "NysR4 (long) protein"		
XX	MO	200159126-72.		
XX	16-AUG-2001.			
XX	08-FEB-2001; 2001WO-GB00509.			
XX	08-FEB-2000; 2000GB-0002840.			
XX	10-APR-2000; 2000GB-0008786.			
XX	14-APR-2000; 2000GB-0009387.			
XX	(UNO-) UNITV NORGES TEKNIKS NATURVITENSKAPELIGE.			
XX	(SMTE) SMITEF STIFTELSEN IND TEK FORSK.			
XX	(ALPH-) ALPHARMA AS.			
XX	(SINV-) SINVENT AS.			
XX	(DZIE/) DZIEGLEWSKA H.			
XX	(ZOTC/) ZOTCHEV S B.			
XX	(SEKU/) SEKUROVA O N.			
XX	(FUAE/) FUAEVIRIK E.			
XX	(BRAU/) BRAUTASER T.			
XX	(STRO/) STROM A R.			
XX	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;			
XX	Valla S, Ellingsen TE, Sletta H, Gulliksen O;			
XX	WPI; 2001-557614/62.			
XX	P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,			
XX	AAE10149, AAE10150.			
XX	New nystatin polyketide synthase polynucleotides and polypeptides,			
XX	useful as antibiotics and antifungals -			
XX	Claim 1; Page 188-254; 266pp; English.			
XX	The present invention relates to the cloning and sequencing of the gene			
XX	cluster encoding a modular type I polyketide synthase (PKS) enzyme			
XX	involved in the biosynthesis of the macrolide antibiotic nystatin.			
XX	The nystatin PKS is useful as antifungal antibiotics. The present			
XX	sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.			
XX	Sequence 125401 BP; 15664 A; 49692 G; 42871 G; 17174 T; 0 other;			
XX	50.7%; Score 51.2; DB 22; Length 125401;			
XX	Query Match			

Best Local Similarity 70.8%; Pred. No. 0.0023;
Matches 68; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 5 GGTGTCGATCGCGCGGTCATGAGCGCGGAGCAGTGTGATCGCGCGGTGAGCAAG 64
DB 13813 GGTGTCGATCGCGCGGTCATGAGCGCGGAGCAGTGTGATCGCGCGGTGAGCAAG 13872
QY 65 GGTGTCGAGCGATCGCGCGGTCGCGGAGCGCGG 100
DB 13873 GGTGTCGAGCGATCGCGCGGTCGCGGAGCGCGG 13908

RESULT 6
AAH79278

ID AAH79278 standard; DNA; 31422 BP.

AC AAH79278;

DT 04-DEC-2001 (first entry)

DE Streptomyces avermiltis coding sequences SEQ ID NO: 2.

KM Avermectin aglycone synthase; AAS; avermectin derivative;

XX drug production; veterinary drug; pesticide; ds.

OS Streptomyces avermiltis.

XX Key Location/Qualifier:

FT CDS 1..14646

FT CDS /tag= a

FT CDS /product= "AAG65266"

FT CDS 14824..31422

FT CDS /tag= b

FT CDS /product= "AAG65267"

XX WO200162939-A1.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-JP01381.

XX 24-FEB-2000; 2000JP-0047405.

XX (KYO) KYOMA HAKKO KOGYO KK.

XX (KITA) KITASATO INST.

XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;

XX WPI; 2001-582053/65.

XX P-PSDB; AAG65266, AAG65267.

XX New modified avermectin aglycone synthase derived from Streptomyces

XX avermiltis used in production of 32,23-dihydroavermectin B1a used in

XX drugs and pesticides -

XX Disclosure; Page 103-149; 257pp; Japanese.

XX The present invention relates to the production of modified derivatives

XX of avermectin aglycone synthase (AAS) derived from Streptomyces

XX avermiltis. The activity of an acyl carrier protein (ACP),

XX beta-ketoadyl carrier protein synthase (KS), acyltransferase (AT),

XX beta-ketoadyl carrier protein reductase (KR), dehydratase (DH), enoyl

XX reductase (ER) and/or thioesterase (TE) domain may be reduced or

XX suppressed. The process can be used in the production of drugs, veterinary

XX drugs and pesticides. The present sequence is a fragment of the S.

XX avermiltis genome.

XX Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

Query Match 50.1%; Score 50.6; DB 22; Length 31422;

Best Local Similarity 70.1%; Pred. to 0.0033; Indels 0; Gaps 0;

Matches 68; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GGTGTCGATCGCGCGGTCATGAGCGCGGAGCAGTGTGATCGCGCGGTGAGCAAG 63
DB 10250 GGTGTCGATCGCGCGGTCATGAGCGCGGAGCAGTGTGATCGCGCGGTGAGCAAG 10309
QY 64 GGTGTCGAGCGATCGCGCGGTCGCGGAGCGCGG 100
DB 10310 GGTGTCGAGCGATCGCGCGGTCGCGGAGCGCGG 10346

RESULT 7
AAH52062

ID AAH52062 standard; DNA; 4851 BP.

AC AAH52062;

DT 04-SEP-2001 (first entry)

DE Mycobacterium tuberculosis potential drug target gene SEQ ID 116.

KM Drug target; growth; organism viability; characterisation; ds.

XX Mycobacterium tuberculosis.

XX WO200135317-A1.

XX 17-MAY-2001.

XX 13-NOV-2000; 2000WO-US31152.

XX 12-NOV-1999; 99US-0165086.

XX 12-NOV-1999; 99US-0165124.

XX 01-FEB-2000; 2000US-0179531.

XX (BEC) UNITV CALIFORNIA.

XX Eisenberg D, Rotstein SH, Marcotte EM;

XX WPI; 2001-329193/34.

XX P-PSDB; AAG81211.

XX Identifying nucleotide or polypeptide sequence for use as drug target,

XX involves providing algorithm that analyzes a functional relationship

XX between nucleotide or polypeptide sequences, and comparing the

XX sequences -

XX Disclosure; Page 131-133; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or

XX polypeptide sequence that may be a drug target, or essential for growth

XX or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

XX represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium

XX tuberculosis proteins which are potential drug targets. The DNA and

XX protein sequences are used to illustrate the method of the invention. The

XX method involves providing an unknown nucleotide or polypeptide sequences,

XX and comparing it to a number of sequences along with at least one

XX algorithm capable of analysing a functional relationship between

XX nucleotide and polypeptide sequences. The method is useful for

XX characterising the function of nucleic acids and polypeptides that may be

XX useful as a target for a drug or essential for the growth or viability of

XX an organism.

XX Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;

Query Match 48.9%; Score 49.4; DB 22; Length 4851;

Best Local Similarity 68.7%; Pred. No. 0.0066;

Matches 68; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Db 744 CGCGCAATGCGATTCGGATCGGTTCCGCCGAGCGG 782

RESULT 8

AAA92302 ID AAA92302 standard; DNA: 31422 BP.

AC AAA92302;

DT 10-JAN-2001 (first entry)

DE 5. avermiltillis avermectin aglycon synthase DNA AEA11 SEQ ID NO:2.

KW Streptomyces avermiltillis: avermectin aglycon synthase; biosynthesis;

KW multifunctional enzyme; polyketide; avermectin; veterinary drug;

KW agrochemical; ds.

OS Streptomyces avermiltillis.

FT Key Location/Qualifiers

FT CDS 1..14646

FT /tag= a

FT /note= "avermectin aglycon synthase protein"

FT CDS 14824..31422

FT /tag= b

FT /note= "avermectin aglycon synthase protein"

XX W0200050605-A1.

XX 31-AUG-2000.

XX 23-FEB-2000; 2000WO-JP01041.

XX 24-FEB-1999; 99JP-0046961.

XX (KITA) KITASATO INST.

XX Omura S, Ikeda H;

XX WPI; 2000-565458/52.

XX P-PSDB; AAB23751, AAB23752.

XX Avermectin aglycone synthase DNA and proteins encoded by all or part of

XX it for the production of avermectin and its derivatives for drug and

XX agrochemical use

XX Claim 2; Page 134-203; 314pp; Japanese.

XX The present sequence represents DNA which encodes avermectin aglycon

XX synthase proteins. Also described are: (1) polypeptides encoded by all

XX or part of the DNA; (2) expression vectors containing the DNA; (3) host

XX cells transformed by the vectors; (4) preparation of the polypeptides

XX by culture of the transformants; (5) preparation of avermectin aglycon

XX or its derivatives by culture of transformed avermectin-producing

XX microorganisms; and (6) oligonucleotides of 5-60 bases in length

XX containing sense or antisense sequences from the avermectin aglycon

XX synthase DNA. The enzymes are useful for the production of modified

XX forms of avermectin and of the intermediates in its biosynthesis, for

XX use as drugs, veterinary drugs and agrochemicals.

XX Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

XX Query Match 48.5%; Score 49; DB 21; Length 31422;

XX Best Local Similarity 69.1%; Pred. No. 0.0077;

XX Matches 67; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

XX Db 10250 GGGGTGTCATGCGCGGTCATGCGCGGAGCAGGTGTCGCGGCGTGAAG 63

XX 4 GGGGTGTCATGCGCGGTCATGCGCGGAGCAGGTGTCGCGGCGTGAAG 63

XX 10250 GGGGTGTCGTCGCGGCTTGAACGACCCCGCTCGTGTCTCCGCGCGGAGCAG 10309

XX 64 CGGTGACAGCGATCGCGGCGGCTTCGCGGCGCGCG 100

XX 10310 AGGTGCGGACATGCGGAGCTGTTCCGCGAGCGCG 10346

RESULT 9

AAT68715 ID AAT68715 standard; DNA: 15872 BP.

AC AAT68715;

DT 01-SEP-1997 (first entry)

DE Streptomyces venezuelae polyketide synthase vep ORF1.

KW Polyketide synthase; polyhydroxyalkanoate monomer synthase;

KW polyhydroxybutyrate; biodegradable polymer; vep gene;

KW metabolic engineering; ss.

OS Streptomyces venezuelae.

FT Key Location/Qualifiers

FT CDS 20..13912

FT /tag= a

FT CDS 14056..14136

FT /tag= b

FT CDS 14148..15827

FT /tag= c

XX W09722711-A1.

XX 26-JUN-1997.

XX 18-DEC-1996; 96WO-US20119.

XX 19-DEC-1995; 95US-0008847.

XX (MINU) UNIV MINNESOTA.

XX Sherman DH, Williams MD, Xue Y;

XX WPI; 1997-341701/31.

XX P-PSDB; AAM19629-30 AND AAM00918.

XX Expression cassettes for production of polyhydroxyalkanoate(s)

XX provide wide range of biodegradable polymers for medical or

XX industrial use

XX Claim 54; Fig 23; 91pp; English.

XX Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide

XX synthase (PKS) gene cluster encoding a polyene of 12 carbons (see

XX also AAM19629-30 and AAM00918). It contains 5 PKS modules, with a 5'

XX loading module and a 3' end domain. Each of the sequenced modules

XX includes a keto-ACP, an acyltransferase, a dehydratase, a keto-

XX reductase and an acyl carrier protein domain. The gene cluster was

XX cloned using a heterologous hybridisation strategy from a genomic

XX DNA library. A novel expression cassette encoding the first module

XX from the vep gene cluster and module 7 from the Streptomyces tlp

XX gene cluster has polyhydroxyalkanoate (PHA) monomer synthase

XX activity and can be used for PHA prodn. in host (esp. insect) cells

XX for use as a biodegradable polymer.

XX Query Match 45.9%; Score 46.4; DB 18; Length 15872;

XX Best Local Similarity 67.7%; Pred. No. 0.031;

XX Matches 65; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

XX Db 10574 GTCTGCGTGGCGGCGGTAAAGGCCCCCGTGGTGTGATCTCCGCGGAGGAGCG 10633

XX 66 GTGACAGCGATCGCGGCGGCTTCGCGGCGCGCGC 101

XX 10634 GTGCGCGAGGCGCGCGCACCTGCGCGAGAGGC 10669

ID	Accession	Standard	DNA	BP
XX	AA687283	standard	DNA	15872 BP
XX	AA687283			
XX	05-JUN-2000	(first entry)		
XX	S. venezuelae vep ORF 1, SEQ ID NO:1.			
XX	Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;			
KW	neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;			
KW	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;			
KW	chronic obstructive pulmonary disease; respiratory inflammation;			
KW	hypercholesterolaemia; crop protection agent; ds.			
XX	Streptomycetes venezuelae ATCC15439.			
OS				
XX	Key	Location/Qualifier		
PH	CDS	20..13912		
FT		/*tag- a		
FT		/product= "vep ORF 1 amino acid sequence #1 (AAV77177)"		
FT		14056..14151		
FT	CDS	/*tag- b		
FT		/product= "vep ORF 1 amino acid sequence #3 (AAV77199)"		
FT		14167..15827		
FT	CDS	/*tag- c		
FT		/product= "vep ORF 1 amino acid sequence #2 (AAV77178)"		
XX				
PN	MO20000620-A2.			
XX				
PD	06-JAN-2000.			
XX				
PF	25-JUN-1999:	99MO-US14398.		
XX				
PR	26-JUN-1998:	98US-0105537.		
XX				
PA	(MINU) UNIV MINNESOTA.			
XX				
PI	Sherman DH, Liu H, Xue Y, Zhao L;			
XX	WPI: 2000-160679/14.			
DR	P-PSDB: AAV77177, AAV77178, AAV77199.			
XX				
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.			
PT	synthesis of methymycin and pikromycin -			
XX				
PS	Example 3; Figure 23; 438pp; English.			
XX				
CC	The invention relates to an isolated and purified nucleic acid segment			
CC	comprising a desosamine biosynthetic gene cluster, a fragment or its			
CC	biologically active variant, where the nucleic acid sequence is not			
CC	derived from the <i>eryC</i> gene cluster of <i>Saccharopolyspora erythraea</i> or			
CC	<i>Streptomyces antibioticus</i> . The invention also relates to a macrolide			
CC	biosynthetic gene cluster, or fragments thereof. The macrolide			
CC	biosynthetic gene cluster encodes proteins which synthesise methymycin,			
CC	pikromycin, neomethymycin, narbomycin or a combination of these			
CC	compounds. Recombinant or augmented cells comprising the desosamine			
CC	and/or macrolide biosynthetic gene clusters are useful for the			
CC	production of biologically active macrolides. The macrolide biosynthetic			
CC	proteins are useful for synthesis of methymycin, pikromycin,			
CC	neomethymycin and narbomycin. The alternative termination of polyketide			
CC	synthesis may be useful to prepare novel antibiotics and			
CC	polyhydroxyalkanoate (PHA) monomers. The compounds produced by the			
CC	recombinant host cells are useful as biopolymers, e.g., in packaging or			
CC	biomedical applications, to engineer PHA monomer synthases or to prepare			
CC	biologically active agents, such as chemotherapeutics.			
CC	Immunosuppressants, agents to treat asthma, chronic obstructive pulmonary			
CC	disease as well as other diseases involving respiratory inflammation,			
CC	cholesterol-lowering agents or macrolide-based antibiotics which are			
CC	active against a variety of organisms, e.g., bacteria, including			

CC	multi-drug resistant pneumococci and other respiratory pathogens, as well
CC	as viral parasitic pathogens, or as crop protection agents (e.g.,
CC	fungicides or insecticides) via expression of polyketides in plants. The
CC	present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC	sequence, designated vep ORF 1 in the specification, which actually
CC	contains 3 open reading frames, which encode proteins AA77177-Y77178 and
CC	AA77199. The vep ORF 1 protein is defined in the specification as a PNA
CC	monomer synthase.
SO	Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;
Query Match	45.9%; Score 46.4; DB 21; Length 15872;
Best Local Similarity	67.7%; Pred. No. 0.031; Mismatches 31; Indels 0; Gaps 0
Matches 65; Conservative	0; Mismatches 31; Indels 0; Gaps 0
Y	6 GTGTGATCGCGCGGTCATAGAGCGCGAGCAGTGATCGCGCGGTGAGCAACG 65
Db	10574 GTCTGCGTCGCGCGGCGGTAACCGCCCGGTCTCGTGTATCTCCGCGCGAGGAAACG 10633
Y	66 GTGCAGCGCATCGCGCGGGTTCGCGAGCGCGCGG 101
Db	10634 GTGGCGCGAGCGCGCGCGCGCAGCTCCGCGAGCAGCGC 10669
RESULT 11	
AA780413	
XX	AA780413 standard; DNA: 43280 BP.
XX	AA780413:
AC	27-FEB-1998 (first entry)
DT	
XX	Tyactone synthase gene cluster.
DE	
XX	Tyactone synthase gene cluster: tylg gene; multifunctional protein;
KW	polyketide; tyactone synthesis; antibiotic; tylosin; ss.
KM	
XX	Streptomyces fradiae.
OS	
XX	
FH	Location/Qualifiers
FT	816..14243
FT	/*tag- a
FT	/transl_except= (pos: 816..818, aa: Met)
FT	/note="ORF1 encodes protein shown in AAW22601"
FT	14351..19945
FT	/*tag- b
FT	/transl_except= (pos: 14351..14353, aa: Met)
FT	/note="ORF2 encodes protein shown in AAW22602"
FT	20010..31199
FT	/*tag- c
FT	/transl_except= (pos: 20010..20012, aa: Met)
FT	/note="ORF3 encodes protein shown in AAW22603"
FT	31232..36067
FT	/*tag- d
FT	/note="ORF4 encodes protein shown in AAW22604"
FT	36249..41774
FT	/*tag- e
FT	/note="ORF5 encodes protein shown in AAW22605"
XX	
XX	EP791655-A2.
XX	
PD	27-AUG-1997.
XX	
PF	19-FEB-1997; 97EP-0301056.
XX	
PR	22-FEB-1996; 96US-0012078.
XX	
PA	(ELIL) LILLY & CO ELI.
XX	
PI	Dehoff BS, Kuhnstoss SA, Rosteck PR, Sutton KL;
DR	WPI: 1997-418046/39.
DR	P-PSDB; AAW22601-W22605.

[illegible]

FT	CDS	57180..59963
ET	/tag= i	
ET	/product= "NysR3 protein"	
ET	60415..61047	
ET	/tag= j	
ET	/product= "NysR4 (short) protein"	
ET	/note= "Cds does not include start codon"	
ET	61736..62497	
ET	/tag= k	
ET	/product= "NysR5 protein"	
ET	/note= "Cds does not include start codon"	
ET	complement (62551..63615)	
ET	/tag= l	
ET	/product= "ORF2 protein"	
ET	/note= "Cds does not include start codon"	
ET	63765..64961	
ET	/tag= m	
ET	/product= "ORF1 protein"	
PX		
PN	WO200159126-A2.	
XX		
PD	16-AUG-2001.	
XX		
PE	08-FEB-2001; 2001WO-GB00509.	
XX		
XX	08-FEB-2000; 2000GB-0002840.	
PR	10-APR-2000; 2000GB-0008786.	
PR	14-APR-2000; 2000GB-0009387.	
XX		
PA	(UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.	
PA	(SNTF) SINTER STIFTTELSEN IND TEK FORSK.	
PA	(ALPH-) ALPHARMA AS.	
PA	(SINV-) SINVENT AS.	
PA	(DZIE/) DZIEGLEMSKA H.	
PA	(ZOTC/) ZOTCHEV S B.	
PA	(SEKU/) SEKUROVA O N.	
PA	(FUAE/) FJAEERYIK E.	
PA	(BRAU/) BRAUTASER T.	
PA	(STRO/) STROM A R.	
PI	Zotchev SB, Sekurova ON, Fjaeryik E, Brautaset T, Strom AR;	
PI	Valla S, Ellingsen TE, Sletta H, Gulliksen O;	
XX		
DR	WP1: 2001-557614/62.	
DR	P-PsDB; AAELI0125, AAELI0126, AAELI0127, AAELI0128, AAELI0129, AAELI0130,	
DR	AAELI0131, AAELI0132, AAELI0133, AAELI0134, AAELI0135, AAELI0136, AAELI0137.	
XX		
PT	New nystatin polyketide synthase polynucleotides and polypeptides,	
PT	useful as antibiotics and antifungals -	
PS		
PS	Claim 2; Page 116-151; 266pp; English.	
CC		
CC	The present invention relates to the cloning and sequencing of the gene	
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme	
CC	involved in the biosynthesis of the macrolide antibiotic nystatin.	
CC	The nystatin PKS is useful as antifungal antibiotics. The present	
CC	sequence is a Streptomyces noursei nysl DNA of nystatin PKS gene cluster.	
XX		
SO	Sequence 65140 BP: 8270 A; 25171 C; 22273 G; 9426 T; 0 other:	
OY		
OY	Query Match	43.8%; Score 44.2; DB 22; Length 65140;
OY	Best Local Similarity	68.5%; Pred.No. 0.092;
OY	Matches	61; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
OY		
OY	6 GTGTGATCGCGCGCTCATATGGCCGAGACAGGTGTCATCGCGGCCTGGACGAAGG 65	
OY		
OY	40799 GTAAGGATCGCGCGCATCAACGGGCCGAACGACACTGTCGTATCGGTGTGAGGATGCC 40858	
OY		
OY	66 GTGCAGGCGGATCGCGGGGGTTCGGCGG 94	
OY		
OY	40859 GCCGTCGAGATCGGGGCGCGGTTCCGCGC 40887	

RESULT 13
AAAI4651/C
ID AAAI4651 standard: DNA; 77536 BP.
XX
AC AAAI4651;
XX
DT 08-AUG-2000 (first entry)
XX
DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
FK-520: polyketide synthase; PKS; gene cluster; immunosuppressant;
KM Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KM FK-506 binding protein; polyketide compound; transplant rejection;
KM graft-versus-host disease; uveitis; alopecia universalis;
KM autoimmune chronic active hepatitis; inflammatory bowel disease;
KM multiple sclerosis; primary biliary cirrhosis; scleroderma;
KM neurite outgrowth; nerve regrowth; Parkinson's disease;
KM Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KM peripheral neuropathy; ss.
XX
OS Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS complement (412..1836)
FT /tag- a
FT /note- "fkpw gene"
FT complement (2020..3579)
FT /tag- b
FT /note- "fkpv gene"
FT 3969..4496
FT /tag- c
FT /note- "fkpr2 gene"
FT complement (4595..5488)
FT /tag- d
FT /note- "fkpr1 gene"
FT 5601..6818
FT /tag- e
FT /note- "fkpe gene"
FT 6808..8052
FT /tag- f
FT /note- "fkpf gene"
FT 8156..8824
FT /tag- g
FT /note- "fkpg gene"
FT complement (9122..9883)
FT /tag- h
FT /note- "fkph gene"
FT complement (9894..10994)
FT /tag- i
FT /note- "fkpi gene"
FT complement (10987..11247)
FT /tag- j
FT /note- "fkpj gene"
FT complement (11244..12092)
FT /tag- k
FT /note- "fkpk gene"
FT complement (12113..13150)
FT /tag- l
FT /note- "fkpl gene"
FT complement (13212..13988)
FT /tag- m
FT /note- "fkpc gene"
FT complement (13452..13662)
FT /tag- n
FT /note- "ACP6"
FT complement (13761..14394)
FT /tag- o
FT /note- "KR6"
FT complement (14517..15294)
FT /tag- p
FT /note- "ER6"
FT complement (15438..16587)
FT /tag- q
FT misc_feature

FT /note- "dehydratase domain (DH) 6"
FT complement (16587..17820)
FT /tag- r
FT /note- "acyltransferase domain (AT) 6"
FT complement (17820..19053)
FT /tag- s
FT /note- "KS6"
FT complement (19116..19326)
FT /tag- t
FT /note- "ACPs"
FT complement (19464..20097)
FT /tag- u
FT /note- "KR5"
FT complement (20241..21420)
FT /tag- v
FT /note- "DH5"
FT complement (21420..22653)
FT /tag- w
FT /note- "AT5"
FT complement (22653..23892)
FT /tag- x
FT /note- "KS5"
FT complement (23992..46573)
FT /tag- y
FT /note- "fkpb gene"
FT complement (24163..24373)
FT /tag- z
FT /note- "ACP4"
FT complement (24997..26146)
FT /tag- aa
FT /note- "DH4 (inactive)"
FT complement (26146..27430)
FT /tag- ab
FT /note- "AT4"
FT complement (27430..28684)
FT /tag- ac
FT /note- "KS4"
FT complement (28750..28960)
FT /tag- ad
FT /note- "ACP3"
FT complement (29092..29740)
FT /tag- ae
FT /note- "KR3"
FT complement (29869..31018)
FT /tag- af
FT /note- "DH3 (inactive)"
FT complement (31018..32185)
FT /tag- ag
FT /note- "AT3"
FT complement (32185..33439)
FT /tag- ah
FT /note- "KS3"
FT complement (33505..33715).
FT /tag- ai
FT /note- "ACP2"
FT complement (33823..34480)
FT /tag- aj
FT /note- "KR2"
FT complement (34606..35749)
FT /tag- ak
FT complement (28750..28960)
FT /tag- al
FT /note- "DH2 (inactive)"
FT complement (35749..37144)
FT /tag- am
FT /note- "AT2"
FT complement (37145..38296)
FT /tag- an
FT /note- "KS2"
FT complement (38371..38581)
FT /tag- ao
FT /note- "ACP1"
FT complement (38677..39307)
FT misc_feature

```

FT      /tag- ap
FT      /note= "KR1"
FT      complement (39442..40609)
FT      /tag- aq
FT      /note= "DH1"
FT      complement (40609..41842)
FT      /tag- ar
FT      /note= "AT1"
FT      complement (41842..43093)
FT      /tag- as
FT      /note= "KS of extender module 1 (KSI)"
FT      complement (43144..43660)
FT      /tag- at
FT      /note= "ACP of loading domain"
FT      complement (43777..44629)
FT      /tag- au
FT      /note= "ER of loading domain"
FT      complement (44974..46573)
FT      /tag- av
FT      /note= "COA ligase of loading domain"
FT      46754..47788
FT      /tag- aw
FT      /note= "Rbo gene"
FT      47785..52272
FT      /tag- ax
FT      /note= "Fkbp gene"
FT      52275..71465
FT      /tag- ay
FT      /note= "Fkba gene"
FT      52362..53576
FT      /tag- az
FT      /note= "KS7"
FT      53577..54716
FT      /tag- bb
FT      /note= "AT7"
FT      54717..55871
FT      /tag- ba
FT      /note= "DH7"
FT      56019..56819
FT      /tag- bc
FT      /note= "EK7"
FT      56943..57575
FT      /tag- bd
FT      /note= "KR7"
FT      57710..57920
FT      /tag- be
FT      /note= "ACP7"
FT      57990..59243
FT      /tag- bf
FT      /note= "KS8"
FT      59244..60398
FT      /tag- bg
FT      /note= "AT8"
FT      60399..61412
FT      /tag- bh
FT      /note= "DH8 (inactive)"

Query Match      43.4%; Score 43.8; DB 21; Length 77536;
Best Local Similarity 69.0%; Pred. No. 0.11;
Matches 60; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY 4 GGGGTGATCGCGGCGGTCGAATGCGCGGAGCGAGTGATGCGCGGCGTGAGCAAG 63
 DB 31536 GCGTGGAGATCGCGCGGTCGACGCGCGGTCGATCGTCTCCGGTGACGAGG 31477
 QY 64 CGGTGACAGCGATCGCGCGGGTTTCG 90
 DB 31476 CGGTGCTGACGTCGCGCGGCTCG 31450

RESULT 14
 AAH51966
 ID AAH51966 standard; DNA: 4209 BP.

```

XX      AC      AAH51966;
XX      XX      04-SEP-2001 (first entry)
XX      DE      Mycobacterium tuberculosis potential drug target gene SEQ ID 20.
XX      KM      Drug target; growth; organism viability; characterisation; ds.
XX      OS      Mycobacterium tuberculosis.
XX      PN      W0200135317-AL.
XX      PD      17-MAY-2001.
XX      PF      13-NOV-2000; 2000MO-US31152.
XX      PR      12-NOV-1999; 99US-0165086.
XX      PR      12-NOV-1999; 99US-0165124.
XX      PR      01-FEB-2000; 2000US-0179531.
XX      PA      (REGC ) UNIV CALIFORNIA.
XX      PI      Eisenberg D, Rotstein SH, Marcotte EM;
XX      DR      WPI: 2001-329193/34.
XX      DR      P-PSDB: AAG81115.
XX      PT      Identifying nucleotide or polypeptide sequence for use as drug target,
XX      PT      involves providing algorithm that analyzes a functional relationship
XX      PT      between nucleotide or polypeptide sequences, and comparing the
XX      PS      sequences -
XX      PS      Disclosure; Page 60-62; 207pp; English.
XX      CC      This invention relates to a method for identifying a nucleotide or
XX      CC      polypeptide sequence that may be a drug target, or essential for growth
XX      CC      or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
XX      CC      represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
XX      CC      tuberculosis proteins which are potential drug targets. The DNA and
XX      CC      protein sequences are used to illustrate the method of the invention. The
XX      CC      method involves providing an unknown nucleotide or polypeptide sequences,
XX      CC      and comparing it to a number of sequences along with at least one
XX      CC      algorithm capable of analysing a functional relationship between
XX      CC      nucleotide and polypeptide sequences. The method is useful for
XX      CC      characterising the function of nucleic acids and polypeptides that may be
XX      CC      useful as a target for a drug or essential for the growth or viability of
XX      CC      an organism.
XX      SQ      Sequence 4209 BP; 751 A; 1067 C; 1412 G; 979 T; 0 other;

Query Match      42.8%; Score 43.2; DB 22; Length 4209;
Best Local Similarity 65.6%; Pred. No. 0.17;
Matches 63; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 6 GTGTGATCGCGGCGGTCGAATGCGCGGAGCGAGTGATGCGCGGCGTGAGCAAGCG 65
 DB 2332 GTGAGCATCGCGCGGTCGAATGTCGCGCTTCGGTGAGTCTGAGCCACGATCG 2391
 QY 66 GTGACGCGATCGCGCGGGTTTCGCGCGCGCGC 101
 DB 2392 GTGAGCGGATCGCTGATCGGCTGCGCGCGCAGGGC 2427

RESULT 15
 AAF31625
 ID AAF31625 standard; DNA: 4209 BP.
 AC AAF31625;
 XX
 XX 09-APR-2001 (first entry)
 DT
 XX Mycobacterium tuberculosis pks (RV0405) gene.

```
XX Mycobacterium tuberculosis; attenuated microorganism;
KW signature tagged transposon mutant; mutant library;
KW mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW vaccine; pks; RV0405; ds.
OS Mycobacterium tuberculosis.
XX
XX WO200102555-A1.
XX
XX 11-JAN-2001.
XX
XX 06-JUL-2000; 2000WO-IB00950.
XX
XX 06-JUL-1999; 98US-0142982.
XX 08-JUL-1999; 98US-0142833.
XX
XX (INSP ) INST PASTEUR.
XX
XX Glacquel B, Gullhot C, Camacho L;
XX
XX WPI: 2001-091804/10.
XX
XX P-PSDB; AAB66459.
XX
XX Screening a mutant library for mutants unable to grow under specific
XX conditions and for identifying loci involved in pathogenicity,
XX comprises using signature tagged transposon mutagenesis
XX
XX Example 8; Page 93-94; 159pp; English.
XX
XX The present sequence is given in a specification relating to a method for
XX screening a library of mutants. The method comprises constructing a
XX library with insertions in genes and/or regulatory regions of the
XX organisms of interest, where the insertion contains a tag and/or a
XX transposon associated with a tag. The mutants are identified by
XX hybridisation of the tags to known sequences. The method is useful for
XX treating an individual suffering from a mycobacterial infection,
XX suspected of being infected with a mycobacterium, or having been
XX exposed to an infectious Mycobacterium. It is also useful for
XX identifying and isolating mutants of actinomycetales and for identifying
XX compounds that have antibiotic activity. The method is used to identify
XX mutants of microorganisms, preferably an actinomycetales, such as
XX M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
XX M. paratuberculosis, that is unable to grow under specific conditions.
XX It is especially useful for identifying loci involved in pathogenicity.
XX It is useful in constructing vaccines. The method can be used to screen
XX multiple libraries concurrently. It can screen libraries of different
XX organisms or different strains of the same organism. The present
XX sequence was isolated from an attenuated mutant of Mycobacterium
XX tuberculosis. The sequence has been disrupted by the insertion of the
XX IS 1096 transposon.
XX
XX Sequence 4209 BP; 751 A; 1067 C; 1412 G; 979 T; 0 other;
XX
XX Query Match 42.8%; Score 43.2; DB 22; Length 4209;
XX Best Local Similarity 65.6%; Pred. No. 0.17;
XX Matches 63; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
XX 6 GTGTGATGCGGGGTCATGAGCGGCGAGTCAGTGTGATGCGGGGCTGGAGCAGGCG 65
XX ||| |||||||||||||||| ||| |||||||||||| ||| |||
XX Db 2332 GTGAGCATGCGGGGCTCATGCTCGGCTTGTGATGATCTTGCTCCACGATGCG 2391
XX ||| |||||||| ||| ||| |||
XX 66 GTGAGGCGATGCGGGGCTTCGCGGCGCGGC 101
XX || |||||||| ||| ||| |||
XX Db 2392 GTGAGCGGATGCGTGTGCTGCGGCGCGAGGC 2427
```

Search completed: November 5, 2002, 13:06:48
Job time : 249.215 secs

SUMMARIES

OM nucleic - nucleic search, using sw model

(without alignments)
6406.881 million cell updates/sec

Title: US-09-724-876-2_COPY_22750_22850

[illegible]

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
```

Not a good idea. Match on

Listing first 45 summaries

Database :

GenEmb1: *

- 1: gb_ba: *
- 2: gb_hlg: *
- 3: gb_in: *
- 4: gb_om: *
- 5: gb_ov: *
- 6: gb_pat: *
- 7: gb_ph: *
- 8: gb_pl: *
- 9: gb_pr: *
- 10: gb_ro: *
- 11: gb_sts: *
- 12: gb_sy: *
- 13: gb_un: *
- 14: gb_vl: *
- 15: em_ba: *
- 16: em_hum: *
- 17: em_hum: *
- 18: em_in: *
- 19: em_mu: *
- 20: em_om: *
- 21: em_ov: *
- 22: em_ov: *
- 23: em_pat: *
- 24: em_ph: *
- 25: em_pl: *
- 26: em_ro: *
- 27: em_sts: *
- 28: em_un: *
- 29: em_vl: *
- 30: em_hlg_hum: *
- 31: em_hlg_in: *
- 32: em_hlg_other: *
- 33: em_hlg_mus: *
- 34: em_hlg_pln: *
- 35: em_hlg_rod: *
- 36: em_hlg_mam: *
- 37: em_hlg_vrt: *
- 38: em_sy: *
- 39: em_hrgo_hum: *
- 40: em_hrgo_mus: *
- 41: em_hrgo_other: *

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF21189	AF21189 Sorangium
2	101	100.0	71989	6	AR172664	AR172664 Sequence
3	99.4	98.4	68750	1	AF210843	AF210843 Sorangium
4	99.4	98.4	68750	6	AR193029	AR193029 Sequence
5	99.4	98.4	68750	6	AR199551	AR199551 Sequence
6	99.4	98.4	68750	6	AR199559	AR199559 Sequence
7	99.4	98.4	68750	6	AR199567	AR199567 Sequence
8	99.4	98.4	68750	6	AR201097	AR201097 Sequence
9	99.4	98.4	68750	6	AR208671	AR208671 Sequence
10	80.8	80.0	16124	1	AX024384	AX024384 Sequence
11	80.8	80.0	16124	6	AX024277	AX024277 Sequence
12	60.2	59.6	20394	1	SNA132222	AU132222 Streptomy
13	60.2	59.6	20394	1	AX067996	AX067996 Sequence
14	60.2	59.6	84985	1	SNA278573	AU278573 Streptomy
15	55.4	54.9	3900	1	AF262754	AF262754 Amycolata
16	53.8	53.3	113193	1	AF357202	AF357202 Streptomy
17	51.2	50.7	27541	6	AX211706	AX211706 Sequence
18	51.2	50.7	123580	1	AF265912	AF265912 Streptomy
19	51.2	50.7	125401	6	AX211739	AX211739 Sequence
20	50.6	50.1	41422	6	E38021	E38021 Aneurectin
21	50.6	50.1	39314	1	SGR300302	SGR300302 Streptomy
22	50.6	50.1	64957	1	AB032367	AB032367 Streptomy
23	50	49.5	78210	1	AB070949	AB070949 Streptomy
24	49.4	48.9	16096	1	AE007124	AE007124 Mycobacte
25	49.4	48.9	24292	1	MTU00024	MTU00024 Mycobacteri
26	49.4	48.9	31175	1	MSGY2	MSGY2 Mycobacteri
27	49.4	48.9	31176	1	MSGY2	MSGY2 Mycobacteri
28	49.2	48.7	26195	1	SCIG7	SCIG7 Mycobacte
29	46.8	46.3	49736	1	AF319998	AF319998 Streptomy
30	46.6	45.1	66808	1	SAU421825	SAU421825 Stigmatel
31	46.4	45.9	17101	1	AB017641	AB017641 Micromono
32	46.4	45.9	43280	1	SFU78289	SFU78289 Streptomyce
33	44.2	43.8	65140	6	AX211705	AX211705 Sequence
34	44	43.6	744050	1	MLEPRN1	AF583917 Mycobacte
35	43.8	43.4	77534	1	AF235504	AF235504 Streptomy
36	43.4	43.0	18023	1	AX024392	AX024392 Sequence
37	43.4	43.0	18023	6	AX024285	AX024285 Sequence
38	43.2	42.8	4209	6	AX069054	AX069054 Sequence
39	43.2	42.8	5544	6	SNA132221	AU132221 Streptomy
40	43.2	42.8	5544	6	AX067994	AX067994 Sequence
41	43.2	42.8	17798	1	AE006546	AE006546 Mycobacte
42	43.2	42.8	22570	1	MTCY4D9	AE8725 Mycobacteri
43	43.2	42.8	84985	1	SNA278573	AU278573 Streptomy
44	43.2	42.8	107379	1	SHGCEPR	X6780 S. hygroscop
45	43	42.6	104326	1	AB070940	AB070940 Streptomy

ALIGNMENTS

RESULT 1				
LOCUS	AF217189			
DEFINITION	AF217189	58733 bp	DNA	linear BCT 09-JUN-2000
	Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.			
ACCESSION	AF217189			
VERSION	AF217189.1	GI:7453554		
KEYWORDS				
SOURCE				
ORGANISM				
	. Polyangium cellulosum Polyangium cellulosum Bacteria: proteobacteria; delta subdivision: Myxobacteria; Myxococcales: Sorangineae; Polyangiaceae; Polyangium. 1 (bases 1 to 58733)			
REFERENCE				

Query Match	Best Local Similarity	Matches 101: Conservative	Score 101: 100.0%	DB 1: 100.0%	Pred. No. 2: 5e-09	Length 58733
QY 1	CGTGGGTGTCATCGCGGCGGTGTCATGAGCGGAGAGGTGATCGCGGCGGTGAGC 60		0	0	0	0
Db 18271	CGTGGGTGTCATCGCGGCGGTGTCATGAGCGGAGAGGTGATCGCGGCGGTGAGC 18330		0	0	0	0
QY 61	AAGCGGTGACGAGCGATCGCGGCGGTGCGCGGCGCGGCG 101		0	0	0	0
Db 18331	AAGCGGTGACGAGCGATCGCGGCGGTGCGCGGCGCGGCG 18371		0	0	0	0
RESULT 2						
LOCUS	ARI172664					
DEFINITION	Sequence 2 from patent US 6303342.					
ACCESSION	ARI172664					
VERSION	ARI172664.1					
KEYWORDS	GI:17912155					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 71989)					
AUTHORS	Jullien,B., Katz,L., Khosla,C. and Tang,L.					
TITLE	Recombinant methods and materials for producing epothilones C and D					
JOURNAL	Patent: US 6303342-A 2 16-OCr-2001;					
FEATURES	Location/Qualifiers					
source	1..71989					
BASE COUNT	10108 a 23531 c 2617 t 11731 t 2 others					

ORIGIN	Query Match	100.0%	Score 101:	DB 6:	Length 71989;
	Best Local Similarity	100.0%	Pred. No. 2.4e-09;		
	Matches 101: Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	CGTGGGTGCATGCGCGCGGTGCATGCGCGGAGCAGGTGATGCGCGCGGTGAGC	60		
Db	18271	CGTGGGTGCATGCGCGCGGTGCATGCGCGGAGCAGGTGATGCGCGCGGTGAGC	18330		
QY	61	AAGCGTGCAGGCGATCGCGCGGGGTTCGCGCGCGCGCGC	101		
Db	18331	AAGCGTGCAGGCGATCGCGCGGGGTTCGCGCGCGCGCGC	18371		
RESULT 3					
LOCUS	AF210843	68750 bp	DNA	linear	BC1 21-JAN-2000
DEFINITION		Sorangium cellulosum strain So ce90 epothilone biosynthesis gene			
ACCESSION	AF210843				
VERSION	AF210843.1	GI:6724237			
KEYWORDS					
SOURCE					
ORGANISM		Polyangium cellulosum.			
		Polyangium cellulosum			
		Bacteria; Proteobacteria; delta subdivision; Myxobacteria;			
		Myxococcales; Sorangineae; Polyangiaceae; Polyangium.			
REFERENCE		1 (bases 1 to 68750)			
AUTHORS		Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milamow, M.,			
		Nowak-Thompson, B., Engel, N., Toupet, C., Strattman, A., Cyr, D.D.,			
		Goriach, J., Mayo, J.M., Hu, A., Golf, S., Schmid, J., and Ligon, J.M.			
		The biosynthetic gene cluster for the microtubule-stabilizing			
		agents epothilones A and B from Sorangium cellulosum So ce90			
		Chem. Biol. 7 (2), 97-109 (2000)			
TITLE		2 (bases 1 to 68750)			
JOURNAL		Direct Submission			
MEDLINE		Molnar, I.			
PUBMED		Submitted (03-DEC-1999) Natural Product Genetics, Novartis			
REFERENCE		Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box			
AUTHORS		12257, Research Triangle Park, NC 27709, USA			
JOURNAL		Location/Qualifiers			
		1. 68750			
FEATURES		/organism="Polyangium cellulosum"			
SOURCE		/strain="So ce90"			
		/db_xref="taxon:56"			
		<1. 1826			
		/note="ORF1"			
		/codon_start=-3			
		/transl_table=11			
		/product="unknown"			
		/protein_id="AAF26904.1"			
		/db_xref="GI:6724238"			
		/translation="ASIDALFARTSARVDDGGRATEKTEHYALEANGIEDLALRBRH			
		LRIOEGSFHMCCLGDLTVELLAHDPLASISPHNARSLRHPDWSDAIIVGPAIV			
		RWIAARCAPGLREYEEERARATQARBLWIAAPCPADLPREDDEANGILPGR			
		MSPEVAERRLRASRYATPELACALIAAGGTGAGPWSGYPAVEMLENTLGGLEPT			
		ALAAASGATSEALRGARLPAFSEVSSKSGOLGIPALMERLRTIYRANGNADN			
		LSRPEREATAAEYRRLRADPPAPPAAGCAIYAGVSSGSLGYDGDALISGDGD			
		IVMOPRISIPVVLACTDPPEFLAPLSMLFYAHNAAGTISKVLTREGSPLYMARN			
		QARMSLVHARGFMAWVNAQMDPDEGAPFVQVRSITMEFHTPPCLHEPAGSAS			
		LACDEEHLWYCELSAGRLLELRHMRHPGAPSRFAYGSEHPIATWYPSLTILATVY			
		WADPDRAIYGVDRKTEVEPIVLAETRHPAHVYSEDRIEALTGPDSDRWVHEHTR			
		SGASTVADYQROLMDRPDKVILNRGLFTTINDRIILILARS"			
		complement(190..3171)			
		/note="ORF2: similar to hypothetical proteins from			
		Mycobacterium tuberculosis and streptomycetes coelcolor and			
		to DP-peptidases"			
		/codon_start=1			
		/transl_table=11			
		/product="unknown"			
		/protein_id="AAF26905.1"			

Best Local Similarity 99.0%; Pred. No. 4.7e-09;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 60
Db 23883 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 23942

Qy 61 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 101
Db 23943 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 23983

RESULT 4
ARI93029
LOCUS ARI93029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source Location/Qualifiers
1. 68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 4.7e-09;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 60
Db 23883 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 23942

Qy 61 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 101
Db 23943 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 23983

RESULT 5
ARI99551
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1. 68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 4.7e-09;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 60
Db 23943 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 23983

Db 23883 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 23942

Qy 61 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 101
Db 23943 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 23983

RESULT 6
ARI99559
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1. 68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 4.7e-09;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 60
Db 23883 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 23942

Qy 61 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 101
Db 23943 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 23983

RESULT 7
ARI99567
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of eptothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1. 68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 4.7e-09;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 60
Db 23883 CGTGGGTGTCATCGCGCGGTCAATGAGCCGAGCAAGTGTGATCGCGGCGTGGAGC 23942

Qy 61 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 101
Db 23943 AAGCGGTGTCATCGCGCGGTTCGCGCGCGCGC 23983

KEYWORDS pimarcin biosynthesis; pims1 gene; polyketide synthase.
SOURCE Streptomyces natalensis.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 20394)
Aparicio,J.F., Collina,A.J., Ceballos,E. and Martin,J.F.
AUTHORS The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimarcin. A new polyketide synthase organization encoded by two subclusters separated by functionalization genes
JOURNAL J. Biol. Chem. 274 (15), 10133-10139 (1999)
MEDLINE 99214571
PUBMED 10187796
REFERENCE 2 (bases 1 to 20394)
Aparicio,J.F.
AUTHORS Direct Submission
TITLE Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOREC, Av. Real, 1-Leon, 24006, SPAIN
JOURNAL Location/Organism
FEATURES
source 1..20394
/organism="Streptomyces natalensis"
/db_xref="taxon:68242"
gene 1..20394
/gene="pims1"
1..20394
/gene="pims1"
/function="pimarcin biosynthesis"
/codon_start=1
/transl_table=1
/product="polyketide synthase"
/protein_id="CAB41041.1"
/db_xref="GI:4678703"
translation="MSNBEKIRELKLRAIDLHETROQIDETEKQREPLAIYSMACR
FPGVKSPEBELMELLRDGVDSFPPRNKRDLDALYHSDAHQSTISYARGGVLLHA
GEFDPGFGGISPREALAMPDQRLLELTAMEAVERAGIDPESLASRTGVEVGHGG
YDAGRRADEVEGHLLTGNHISIASGRISVYLDEGLPDLVDPAACSSLYALILAH
ALRDECMALVGATVMTSPFOMEFERSROGLADPGRCFPAADGDTGSEBGLL
LVERLDAVNRGYPVLAVIKGSANODGASNGLTPASQOORVROALTGAGIAAD
IDAEAGCTGTTCGDPYBAHALLATYGOORAADPCCIGSKSNIGHTQAAGIAGYA
KMLAMRHGILPRTLHIDEPITGHVDMSEGNRLAEPEWPSAGRPRAAVSFGISG
TNAHVILEQAFAHEAEPAPPAARPGALPWLTARTAGLRAQADRILGRHLRADAE
PAAVAHALDRTITMEHRAVVAAGDREELTLDALAGRTANGVSVKASATL
FAGGSGORPGKRELLHAHPYATAFVACAEIDPHDLRLDIPFAEGESGALAD
OTATYOAALFLETFALFRIVESMGVAPFAVGSIGELTAAHVSGLTLIDHAALVA
RGTMOALPAGGAMVAVOATEDETREBLAGHEDHVALAANGPSTYISGDEQAVTEL
AAHAEQGRKTKRLVSHRSPHMDLMEDFRVARGLTFHARITVSVSTVTLAEP
EDELRSPTYWROQRETVRCFAVRTLEAGVTVEELGTGGVLTPIWVODCLTTLLEP
VLPHLRTGREYVALTEGATVAFVAGVPVDRSAFPGAPGTSRADLPTIYAFQRMWYL
DPADHGEAAALAAAGAGAEFWAVEREDLOELSAVLAIDGSEADSIGSEFTLTSWNR
ORTTOAADRESYRTHMAPTASGGPTATGHWIYVLPGGTDDPWTARILDALINDGL
HTDRELPADHEPDMARKHPYDVYLCILADDERPTSCPTPRGLATITDAANRPECA
GICAPLWCVTTRGAIVDRHEALKSPDLQATMGIDKRVAALESPOSWGGILTLIPNDLGE
AVSALSTLAGEEDQVAVRPAGVFAARRLERITPGGDYGRKSTHGTVALYVGGTGLAL
HLAHMLADAGAEHLVLTGRGPOAPGAPBELLAALATLDGKAVTTLACDAADDAALVAL
ADIPHLPLTGVNAAGVLDGVDLALTPEFETVLRPKKAAQNLEHILJODLDOHF
VLFSSIVGIGNAGQANAANAYLALDAHERLAOGGLPATSVSMGPGQAAWMSDAA
DRMSRGLLPMMAAPRRRRAPALAQCMGTQVYADIDMSAYAPALTITAYRPSLTGDIIE
ARRALPABEGPREBRSPULDRIGALPPABDEKAPLTWREBAALVLIHPSIDVDAOR
AFRQGFDSLMAVDLNRRLSATGLRLPATLLFDPPLLAACLRSEVLEGAAPAVY
VOASTALDEVALIIGMACRPPGGVSHPEALMRLAEGDALTTPMAPRDLTDLRH
PDPHOGTSYARGGGFLDGAADPADDFGSGISPREALAMPDQRLLELTWELDEAGID
PELSRSGSTGVFAGTNTODGTALDAADODAGGRLTGNMANSVSGSVSTPFGEGCA
LYVDTACSSSTIVALHMAAOALROEGCSLAAGCTYVATPSSSYEPERORGLADGKC
KPPRAADGTGMBGVGLLVERLSDKRNGHOVLAVRGSANQDASNGLSAPSC
SQOQVITQOALANARVAASEVDVAHGTGTLTGPTEAOALVAYGGERPLTGAVS
NGHITQOAGVAGVAKMVLAMRHGILPRTLHIDEPITGHVDMSEGNRLAEPEWPS
GHPRAVSAFEGISGTNAHVILEPAEOLPLVEBPSAEPDAPATADRPITADGGA
PLISASSESALRQOARLHSHLERDAPRLTDAAYTLMHRTAFARAVRADADRA
ALATLTLAGEADRAVDTGTAHTGRDAVLFSSGGSORIGMRSLSTRYPRAPAEPT
VCAALDHLRPLRDLVYRGDEBELTITVAOGLFAIEVALFRLVSWGRPHYVAG
HSVGEILAAHVAGVFLADACALVAANGRTAOLPAGAMAAIRATEDVELPHLADV
STAAVNGPSSVYVSGAEHVAISIAHPEGAGRKTTRLRVSASHISPLMDLDFRAV
AEGILTYGEPELAVVSVNTGOLATPDQLRTPEYVWTHVRAAVRFADIRALGAEVTR


```

/product="hypothetical protein"
/protein_id="CAC20917.1"
/db_xref="GI:12035068"
/translation="RRPDAFREATEPYGALLAPIMPERSADRVGEKTKWTSWDSQARFL
AHDACEVLTLACGHPPLVLEDDLOWMDPTASLDLRLSLRCCGHGLGVLTPARKKE
TESDPTLRIYSDVLRGPRATRLRCLGCPAPVAGVALVNAHGSVDEVEYLRRCSE
GMPYVNMOLLISVGBARRLRPDAVDALLALIPGVGALEAMNORPAALPEVYLIRRE
AVTAGEVDTLRLRATCTDASVYNGLEBSAITAGLIGEDPHHRCGLHNAHMLNDETLFD
ELPDRRRRLIARAKADALIYARRKGVGEETLERYAHNSHMAKSLPAEYVLPRLILAA
EHAEOBAMVEOYEIVLARAVALVGFLLPDDPATVETLROOKLILQGVGLATVRGGLHE
AETALTRRALSTVYHADEPDSVMALACASLIVGRYDASRGLSLRLDLSWQTRREVY
AVLGAAYEGEIVLYHVRGMLPOLAAELKEGVSRADHPNEGHARLTPODHPRVSCRSY
NFTFVNLGHDGIDARRELSTLEYEERSDSSPALVYDGVGAALAEAGDQRTPLALASS
DEGVVAADHGLITGKAMLVGVKMGKGLVNAHDEGRLALIMTSLIAELGDSRTYLRHLI
HGLIGLQAHNHNGOLEEARTTFRMLLAAVEQRREYVILHPLALPTPLIHETLGHGDE
ALADDDPGRGV"
2190..3566
/gene="p1mk"
2190..3566
/CDs

```

gen
CDS

SMUGLVKAGARAVNTTIGETAVSLADRLAARPEREOTKNDVPTREAAVVLGHASS
ONVDAREFEELGDFSLTSEVELNRLSATVTEGRSLATVVEFDSKNDPRDLAARVADLAA
QSLADGSAQAGFLPYAAPTETDSLEMPIDALDSCGVPEAORMLSLAGLPSEPTA
ELEDDLPATLAEAGPCTPRLICSTPTAGVYEVRIARASPERGPRVSALEPAGAT
GPELPATTAAYAVIAEALRASDNMPVLVGHSSGAFAYLAALILEMTWIGREPAY
VLDDITSLRHEONESIDAGLMRHEPVDVSVPTNTNLSLMAHMKMLNDLEVRH
TTPVLTIRAKTEFTGIDTGTIGEDHSPVDVSDADHSMVRKDAPEARIVRE
WLDLSGNA

gene
complement(9772, 15198)
/gene="pims3"
complement(9772, 15198)
/gene="pims3"
/function="polyketide synthase"

transl_table=1
/product="pims3 protein"
/protein_id="CAC26920.1"
/db_xref="GI:1205071"

/translation="MTTSETVVAALASLKEAENLRQNRKLVAAETIATVAMAC
RTPGVQSPEDLWELLTNGKDAIGAPTRGMDDLADGADRNNAVSOEGFLDG
VGFADPFGISRESVYMDPORLLETSWEAIEBAGIDPTLRSGVFGTNGO
DYAYLLVRLADATGDTGTAASATSGRLTYLGEPAVTVDTACSSSLVALTAS
HALRAGECSLALGVNMSPGSLMEFRGGLADGCKAFSDTADGTGMEGVY
LYVERLSDARNRHPVLAIVRGSAVNOGDSNGETAPSGPOORVIROLAAGSAS
DVVVAHGTGTPIDPIEAOALATYGGDROPERLLIGVSNNGHQAAGAS
IKITLMBRGVLPKTLVADVPSHVDMSGEVELLEPEMBKPADPMHAGSREIS
GNANHYVQAEDEEPVNEPGLAPVYVWLVSAKTEELNAQALATISSRGTASPD
VPSLSAGSAFEHRVLAQCPDGLSENARATSTRPLAVLSEGCAGORAGRE
LVKRFPEALDEVLAFHFDAGLRDWMFADBGDLDTGTOPALEIIVALLRLASL
GVARDFVGHSIGELIAAHVAGVSLDCAVALAARALMAQAPAGAVAAVQASSE
VVARLDVGSIAAVNPEAVVLADGETAVLRIEAAAGRGKPLVSHAFESPLME
PLEDERRVAEGLSTGAPJIPLVSNVTGELVAEELCSPEYVVRHRETRFEDGSA
LEAEASAFLELGPDPVLTAMARHTLEKAVLVKRPKEPEETALLTALHNVAGI
GVWMSGIFAGTGARVLDJTYPOHERYPRPAVLVSGDVAAGLRAEHPILGATLTL
AESGEVLTGRLSLQTHPNLTDHAGNENFPEPTGTELEMAHADGDDREVDLTM
FGSTVAMPDQAVAVDLEGVDEAGTGTTRTILHAREDDPROMHQAQGTITGSEERV
FGLHALLDVAVLOSRIHNVGSADDOELFANRGVSLHAGSAYLVLRVYKGDMS
IAVNDABGVLSAEAMVIRATRDPTGGAHNGERBELLSLEWSEGPPEPPACV
CSVLGDELGVASVSLDLSGDEDFVLPVSGVDPAAHHEITARTLLDLOWS
GRLGSSGRLVETRNAVIADEGESVADVAANAAGLVSAQSESGRIALVDLEDS
GDLAEVLALPGLLAGDIAQFVREGAVNAGLVESNEGAGVSPRMDADGVLYVGC
TGLGRELARHLEVERGVHILVSRSGEPAGVSGLSBELAHAGVEVVRACDVADR
AAVHALVAAVPAHEHPLTAVIHTAGVLDGLVLSLPERLSTAPRVDAMNHEATK
DLDLAFLVYSSISGVIGPGQANVAGNVFDALAMHMGGLPACOSLAMAMTQDG
GMTATLSDTDMQMASYGNAPLSRORGALFPAACRSNAHLVAVRGVGPASAPV
PSYRLGVKAGARTAVNTJGETRAVSLADRLAARPEREPTLDTVIRTAAVLGHAS
AKNVDAAREFENLGDPSLIVELRNLSLTLGLRSLAALVPPYNPRTALADHVSILY
ERATGPALLAELERDSILANGAPARTRAVNTANLADILDDMRDPAEDGCAEVAE
RLBASTDIEAFIDELRHSR

gene
complement(15210, 43733)
/gene="pims2"
complement(15210, 43733)
/gene="pims2"
/function="polyketide synthase"

transl_table=1
/product="pims2 protein"
/protein_id="CAC26921.1"
/db_xref="GI:1205072"

/translation="MDOEQLLRDYLRASADLRKRNORGELEAARPEITAYVMSKR
FPGVRSPELWSLADAGGILKSPADGWELADAGESSGFLHDAEFDASFG
ISPRELALDPOORLLESWEAFERAGIDPAVARTGMFGVAGAMPQRYRGPDVY
OGFALGTTTYSVIGRLAYFGAVPAVTVDTACSSSLVALHMSLQSGCSLDA
AGVWMSPTTEVEFNQGLSADGKRSFADSGTGMSEGVYLVLESEARNG
HEILAVIRGSANODASNLITAPDSRSQRIEQLVARSALADVEDVVEHNGCTT
LGDPEAOALATYGGDIDRPLLVGLSVSNLGHQAAGVYKMLARHGLV
RTLHVDAPESSHVDMOGAVILITLLEHPVPMQSGPRRAVSGFSGTNGTILDEEPA
APSPGGEEDTAPVAEPLAAGVPMVLSGRTREALBAOARLLEHLSPEGLADVA
YSLATRSGLIHRAAFTAGPGAGVATLADAGLGNODPAPRTAKFLIRAGC
GSGORMGRELAARPPVPAALDEVLGHHDECLERLKVLRABSTPPAALLDGTG
AOPALPRAIVALTAESERTITPDLAGHSIGELIAAHVAGVFTTLADADVALVAREH
MOLPBGAMVSLAEETLPLLEDRPRLSTIAAVNPAVVIAGEADVVAVAHEF

Query Match 59.6%; Score 60.2; DB 1; Length 84985;

Best Local Similarity 76.3%; Pred. No. 0.037;

Matches 74; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 4 GGGTGTGATCGGGGGGTCATAGGCGCGGAGCGGTGGATCGGCGCGGACGACG 63
Db 69235 GCCTTCGCTCGCGCGGCGGCTCAATGCGCGTCTCGGTCTGATCTCGGCGCGGAGACG 69176
Qy 64 CGGTGACGCGGATCGGCGGCGGCTTCCGCGCGCGG 100
Db 69175 CGGTGCTGCGCATCGGCGCGGCGGCGGAGG 69139

RESULT 15
AF262754
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AF262754 3900 bp DNA linear BCT 27-SEP-2000
Amycolatopsis mediterranei DSM 46095 rifamycin biosynthetic gene
cluster, partial sequence.
AF262754.1 GI:10312147
Amycolatopsis mediterranei.
Amycolatopsis mediterranei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;
Amycolatopsis.
1 (bases 1 to 3900)
Lal, R. and Kaur, H.
Cloning and partial characterization of presumptive rifamycin
biosynthetic gene cluster from the actinomycete Amycolatopsis
mediterranei DSM 46095
Microbiol. Res. (2000) In press
2 (bases 1 to 3900)
Lal, R. and Kaur, H.
Direct Submission
Submitted (02-MAY-2000) Zoology, University of Delhi, Delhi 110007,
India
Location/Qualifiers
1. 3900
/organism="Amycolatopsis mediterranei"
/strain="DSM 46095"
/db_xref="taxon:33910"
106..>3900
/note="similar to Amycolatopsis mediterranei S699
rifamycin biosynthetic gene cluster; contains module 2
ACP, all of module 3, including KS, AT, and ACP, as well as
part of module 4"
/codon_start=1
/transl_table=1
/product="rifamycin polyketide synthase"
/protein_id="AF262754.1"
/db_xref="GI:10312148"
/translation="MIFDYPKPGADLHRLGRLEGNASRAEITGTAANDPEIATYA
MACRPPGVGHPSEDIMRLVADGVDAVTEFPATRGWDDRLYHEDDHEGTYVHGAPE

Lawrence Berkeley National Lab
One Cyclotron Rd., Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic sequence AC005554
plate: 37 row: G column: 5
High quality sequence stop: 514.
Location/Qualifiers

QY 61 CGCGGCGCGGAGGCTGCGCGTGTGTG 91
 DB 103 CGCGGCGAGGAGCCCTCGCCGAGCTGTG 133

RESULT 4
BM488815
LOCUS 597 bp mRNA linear EST 07-FEB-2002
DEFINITION pgm2n.pk008.p20 Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA clone pgm2n.pk008.p20 5' similar to ref|NP_446119.1 (NM_053667) growth suppressor 1; leprecan [Rattus norvegicus]
 gb|AAD51875.1|AF087433.1 (AF087433) leprecan [Rattus norvegicus], mRNA sequence.

ACCESSION BM488815
VERSION BM488815.1 GI:18609746
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus

REFERENCE
AUTHORS 1 (bases 1 to 597)
TITLE Esrs from Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library, USDA/IFARS Animal Genome Project

JOURNAL Unpublished (2002)
COMMENT Contact: Larry A. Coghurn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: coghurnudel.edu, www.chickest.udel.edu.

FEATURES
source 1..597
 location/Qualifiers
 /organism="Gallus gallus"
 /strain="Commercial broiler and Ottawa Res. Centre Strains 90 & 21"
 /db_xref="taxon:9031"
 /clone="pgm2n.pk008.p20"
 /clone.lib="Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n)"
 /sex="Male and Female"
 /tissue_type="Breast muscle, leg muscle and epiphyseal growth plate"
 /dev_stage="Breast, leg: Embryo(d19); post-hatch(1d,1.3,5,7,9,11 weeks); growth plate(1d,7d,14d post-hatch)"
 /lab_host="E. coli EMDH10B"
 /note="vector: pCMVSPORT6; library made from equivalent pools of total RNA isolated from each tissue (embryonic muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth plate 33.3% of the final RNA pool). Single pass sequencing from 5'-end"

BASE COUNT 78 a 200 c 199 g 90 t 30 others
ORIGIN

Query Match 35.2%; Score 35.6; DB 13; Length 597;
Best Local Similarity 60.2%; Pred. No. 77;
Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 4 GCACGGCTTTGGCCCGGCGGCGGCAAGACGTTCTGCGCGAGCGGCGGCTACGC 63
 DB 68 GCCCCTCTGCGCGGAGACCCCGGACGCGCTTCGTCGCGGCGGCGGAGCGGCTACGC 127

QY 64 GCGGCGGAGGCTGCGCGTGTGTGCTCAAGCGC 101
 DB 128 GCGGGGAGCTGCGCCCGCTGTGCTGCAATGAGC 165

RESULT 5
AG159162/c

LOCUS AG159162 885 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-024J17.TJ, genomic survey sequence.

ACCESSION AG159162
VERSION AG159162.1 GI:16688840
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA, clone.lib:RP43-024J17.TJ.

ORGANISM Pan troglodytes
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Torok,i.Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library RP43-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 885)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Torok,i.Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan 1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpes@qsc.riken.go.jp, URL:http://hgp-qsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RP43-43 This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.

PRIMERS
 Sequencing: TJ
LIBRARY Vector : pBACe3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI.
FEATURES
source 1..885
 location/Qualifiers
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-024J17.TJ"
 /sex="male"
 /cell_type="lymphocytes"
 /clone.lib="RP43-43 Chimpanzee Male BAC Library"

BASE COUNT 14 a 473 c 357 g 16 t 25 others
ORIGIN

Query Match 34.3%; Score 34.6; DB 17; Length 885;
Best Local Similarity 62.7%; Pred. No. 1.2e+02;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 4 GCACGGCTTTGGCCCGGCGGCGGCAAGACGTTCTGCGCGAGCGGCGGCTACGC 63
 DB 565 GCGCGGCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 506

QY 64 GCGGCGGAGGCTGCGCGTGTGTGCTCAAGCGC 86
 DB 505 GGGCGCGCGGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 483

RESULT 6
CNS00720
LOCUS 932 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14809 of RP43-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL066742
VERSION AL066742.1 GI:4945205
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 932)

AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 131 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org/TheBDGP/Drosophila>
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RP11-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
SOURCE 1. 932
Location/Qualifiers
/db_xref="taxon:7217"
/clone_lib="BACRI4B09"
/clone_1lb="RPCI-911"
/note="end : T7"

BASE COUNT 155 a 202 c 241 g 91 t 243 others
ORIGIN

Query Match 33.9% Score 34.2; DB 17; Length 932;
Best Local Similarity 46.4%; Pred. No. 1.5e+02;
Matches 45; Conservative 17; Mismatches 35; Indels 0; Gaps 0;

OY 5 CACGCGCTTTCGCCCGCGCGGCGTGCACAGCTTCGCGCGCGAGCGGCGCTACGCG 64
Db 692 CSSCSCSCCG 751
OY 65 CGGCGCGAGGCGTGCCTGCGTGTGCTC AAGCGCG 101
Db 752 SSSCGCGSGCG 788

RESULT 7
BQ674384 934 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8211307 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6276688
DEFINITION 5', mRNA sequence.
ACCESSION BQ674384
VERSION BQ674384.1 GI:21785218
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 934)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: crabs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Rb1a Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNCM2461 row: n column: 17
High quality sequence stop: 464.
Location/Qualifiers

FEATURES
SOURCE 1. 934
Location/Qualifiers
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone_image="6276688"
/clone_1lb="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: salivary gland; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACTG(6). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 125 a 361 c 347 g 96 t 5 others
ORIGIN
Query Match 33.9% Score 34.2; DB 14; Length 934;
Best Local Similarity 60.7%; Pred. No. 1.5e+02;
Matches 54; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 3 TGCACGCGCTTTCGCCCGCGCGGCGTGCACAGCTTCGCGCGCGAGCGGCGCTACG 62
Db 526 TGCCTGCG 585
OY 63 CGCGCGCGCGAGCGCTGCGCGCGCGTGTGCGT 91
Db 586 CCG 614

RESULT 8
BQ805311 715 bp mRNA linear EST 31-JUL-2002
LOCUS WHE3565.D05_H092S Wheat developing grains cDNA library Triticum
DEFINITION aestivum cDNA clone WHE3565.D05_H09, mRNA sequence.
ACCESSION BQ805311
VERSION BQ805311.1 GI:22029520
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.
1 (bases 1 to 715)
AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin
,K., Crossman,C., Fenton,R.D., Iazo,G.R., Pham,J., Rausch,C.J.,
Wilson,C. and Woo,J.
TITLE The structure and function of the expressed portion of the wheat
genomes - Developing grains cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: anderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES
SOURCE 1. 715
Location/Qualifiers

/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3565.D05_H09"
/clone_1lb="Wheat developing grains cDNA library"
/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/note="Vector: Lambda ZAP II, excised phagemid; Site_1:
EcoRI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
24oc/17oc day/night, well-watered, with post-anthesis

RESULT 11			
BI869677			
LOCUS	640 bp	mRNA	linear
BI869677			EST 11-OCT-2001

[illegible][illegible]

Wed Nov 6 09:54:34 2002

us-09-724-876-2_copy_21350_21450.rst

Page 8

Db 373 NCCCCGCCCCGCCCCGNNCGGNNNNNN:NC 343

Search completed: November 5, 2002, 16:17:53
Job time : 1120.44 secs

APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 4689
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-34

Query Match 41.6%; Score 42; DB 10; Length 4689;
Best Local Similarity 64.3%; Pred. No. 0.0032;
Matches 63; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 4 GCACGCGCTTCGCCCGGCGGCGGTGCAAGACGTTCTCGCGCGGCGGAGCGGCTACGC 63
DB 753 GCGCGGCGCTTGCGCGCGGAGCGCGCGGTGCAAGCGCTTCGCCACCTGCGGAGCGGCTTGG 812
QY 64 GCGGCGCGGAGCGCGCGCGGTGCTGCTCAAGCGC 101
DB 813 CCGCGCGGAGCGCGCGGAGTCTGCTGAGCGCGC 850

RESULT 7
US-09-924-256A-69
Sequence 69, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924,256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 69
LENGTH: 658
TYPE: DNA
ORGANISM: Usnea florida
US-09-924-256A-69

Query Match 41.4%; Score 41.8; DB 10; Length 658;
Best Local Similarity 63.4%; Pred. No. 0.004;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 GATGACCGCGCTTCGCCCGGCGGCGGTGCAAGACGTTCTCGCGCGGAGCGGAGCGCTA 60
DB 367 GCTGAATGCTTCTCTACGCGCGCGCTCCCGCATGTGGAGTGGGAGCGGAGGCTA 426
QY 61 CGCGCGCGCGGAGCGCGCGGTGCTGCTCAAGCGC 101
DB 427 CGCTGCTGGGAGCGCGGTGCTGTAGTGTGTAAGCGC 467

RESULT 8
US-09-924-256A-53
Sequence 53, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap

APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924,256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 53
LENGTH: 643
TYPE: DNA
ORGANISM: Peltiligera neopolydactyla
US-09-924-256A-53

Query Match 38.4%; Score 38.8; DB 10; Length 643;
Best Local Similarity 67.1%; Pred. No. 0.024;
Matches 55; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 18 CCGGCGCGCGGTGCAAGACGTTCTCGCGCGGAGCGGCTACGCGCGGCGGCGGCT 77
DB 364 CTGATGCTGCTTGCAAACTTGTGATGCTGATGCTGATGCTGCGGCGGAGGTT 443
QY 78 GCGCGCGGTGCTGCTCAAGCG 99
DB 444 GTGCTGTGTATTTCTCAAGC 465

RESULT 9
US-09-924-256A-29
Sequence 29, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Waters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924,256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 29
LENGTH: 643
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Clone
OTHER INFORMATION: Xa.cum.6H
US-09-924-256A-29

Query Match 32.1%; Score 32.4; DB 10; Length 643;
Best Local Similarity 62.2%; Pred. No. 1.1;
Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 17 CCGGCGCGCGGTGCAAGACGTTCTCGCGCGGAGCGGCTACGCGGCGGCGGAGGCT 76
DB 383 CTTAGCGGCGCATGCAAGCGTTTGTATGCTTCTGCTGATGCTACTGCGGAGAGAGA 442
QY 77 TCGCGCGGTGCTGCTCAAGC 98
DB 443 GTTGTGCTGCTGCTTAAGC 464

RESULT 10
US-09-764-855-227
Sequence 227, Application US/09764855
Patent No. US20020119919A1

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult P110 or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 227
; LENGTH: 3966
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-855-227
```

```

Query Match          31.7%; Score 32; DB 10; Length 3966;
Best Local Similarity 58.3%; Pred No. 1.2;
Matches 56; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
```

```

QY 2 ATGCACGCGCTTTCGCCCGCGCGGTCACAGAGCTTTCGCCGACGCGAGCGCTAC 61
Db 3802 AGAACCGCGCTCCGCTCAGCGTGGCCCKCAGACACACAGCGCGGAGTGACCTGTC 3861
QY 62 GCGCGGCGCGAGGCGCTCCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 97
Db 3862 GTGCAGAGCTGTGAAGCAGCGCGGTGGCGTGCTGAG 3897
```

```

RESULT 11
US-09-861-289-19
; Sequence 19, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 1281
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-19
```

```

Query Match          30.5%; Score 30.8; DB 10; Length 1281;
Best Local Similarity 57.1%; Pred. No. 2.7;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```

QY 1 GATGCACGCGCTTTCGCCCGCGCGGTCCAGACGTTTCGCCCGCGAGCGGCTA 60
Db 924 GATGCACGCGCTTTCGCCCGCGGCTGCTGCGGATCATCCACACGCGCGGCGACCTA 983
QY 61 CGCGCGGCGGAGGCGTCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 98
Db 984 CGCGACCGCGGTGTATCAACGCGGTGCGCGCAGGTCTATGC 1021
```

```

RESULT 12
US-09-861-289-3
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
```

```

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-3
```

```

Query Match          30.5%; Score 30.8; DB 10; Length 13613;
Best Local Similarity 57.1%; Pred. No. 2.3;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
```

```

QY 1 GATGCACGCGCTTTCGCCCGCGCGGTCCAGACGTTTCGCCGCGAGCGGAGCGCTA 60
Db 3085 GATGCACGCGCTTTCGCCGAGCTGCTGCGGATCATCCACACGCGCGGCGGACCTA 3144
QY 61 CGCGCGGCGGAGGCGCTCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 98
Db 3145 CGCGACCGCGGTGTATCAACGCGGTGCGCGAGGTCTATGC 3182
```

```

RESULT 13
US-09-805-177-1
; Sequence 1, Application US/09805177
; Patent No. US2002001805A1
; GENERAL INFORMATION:
; APPLICANT: Richard Bruce Roden
; APPLICANT: Honami Nozaki
; TITLE OF INVENTION: IMMUNOGENIC OVARIAN CANCER GENES
; FILE REFERENCE: 031787.0090
; CURRENT APPLICATION NUMBER: US/09/805,177
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/189,226
; PRIOR FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: 60/258,452
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(654)
US-09-805-177-1
```

```

Query Match          30.1%; Score 30.4; DB 10; Length 753;
Best Local Similarity 61.2%; Pred. No. 3.5;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```

QY 1 GATGCACGCGCTTTCGCCCGCGCGGTCCAGACGTTTCGCCGCGAGCGGAGCGCTA 60
Db 162 GATGCACGCGCTTTCGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 221
QY 61 CGCGCGGCGGAGGCGCTGCG 80
Db 222 CGCGCGCGGCTATGCGCTCG 241
```

```

RESULT 14
US-09-924-256A-79
; Sequence 79, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Maao, Vivian
```



```
RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409
```

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/3:5,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match
Best Local Similarity 98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.6e-16;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATGCACGGCGCTTTCGCCCGGGGGGGGTCAGAGACGTTCTCGCCGAGCGGAGCGGCTA 60
DB 26962 GATGCACGACACTTTCGCCCGGGGGGGGTGCAGACGTTCTCGCCGAGCGGAGCGGCTA 27021
QY 61 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 101
DB 27022 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 27062
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match
Best Local Similarity 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.6e-16;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGCACGGCGCTTTCGCCCGGGGGGGTGCAGAGACGTTCTCGCCGAGCGGAGCGGCTA 60
DB 26962 GATGCACGACACTTTCGCCCGGGGGGGGTGCAGACGTTCTCGCCGAGCGGAGCGGCTA 27021
QY 61 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 101
DB 27022 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 27062
```

RESULT 4
US-09-567-969-1

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match
Best Local Similarity 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.6e-16;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGCACGGCGCTTTCGCCCGGGGGGGTGCAGAGACGTTCTCGCCGAGCGGAGCGGCTA 60
DB 26962 GATGCACGACACTTTCGCCCGGGGGGGTGCAGAGACGTTCTCGCCGAGCGGAGCGGCTA 27021
QY 61 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 101
DB 27022 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 27062
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match
Best Local Similarity 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.6e-16;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGCACGGCGCTTTCGCCCGGGGGGGTGCAGAGACGTTCTCGCCGAGCGGAGCGGCTA 60
DB 26962 GATGCACGACACTTTCGCCCGGGGGGGTGCAGAGACGTTCTCGCCGAGCGGAGCGGCTA 27021
QY 61 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 101
DB 27022 CGCGGGGGCCGAGGGGCTGCGCCGCTGTGTGCTCAAGCGGC 27062
```

```
RESULT 6
US-09-568-486-1
: Sequence 1, Application US/09568486
: Patent No. 6353459.
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,486
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.6e-16;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Yy 1 GATGACACGCGCTTTCGCCCGGGGGGGGTGCAAGACGTTTCGGCCGACGGAGCGGCTA 60
Db 26962 GATGACACCACTTTCGCCCGGGGGGGGTGCAAGACGTTTCGGCCGACGGAGCGGCTA 27021

Yy 61 CGCGGGGGCGAGGGCGCTGCGCTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGGCGAGGGCGCTGCGCTGTGTGCTCAAGCGGC 27062

RESULT 7
US-09-568-472-1
: Sequence 1, Application US/09568472
: Patent No. 6358719
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/568,472
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.6e-16;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Yy 1 GATGACACGCGCTTTCGCCCGGGGGGGGTGCAAGACGTTTCGGCCGACGGAGCGGCTA 60
Db 26962 GATGACACCACTTTCGCCCGGGGGGGGTGCAAGACGTTTCGGCCGACGGAGCGGCTA 27021

Yy 61 CGCGGGGGCGAGGGCGCTGCGCTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGGCGAGGGCGCTGCGCTGTGTGCTCAAGCGGC 27062

RESULT 8
US-09-567-899-1
: Sequence 1, Application US/09567899
: Patent No. 6383787
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,899
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 3.6e-16;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Yy 1 GATGACACGCGCTTTCGCCCGGGGGGGGTGCAAGACGTTTCGGCCGACGGAGCGGCTA 60
Db 26962 GATGACACCACTTTCGCCCGGGGGGGGTGCAAGACGTTTCGGCCGACGGAGCGGCTA 27021

Yy 61 CGCGGGGGCGAGGGCGCTGCGCTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGGCGAGGGCGCTGCGCTGTGTGCTCAAGCGGC 27062

RESULT 9
US-09-103-840A-2/c
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patent Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          62.0%; Score 62.6; DB 4; Length 4403765;
Best Local Similarity 79.6%; Pred. No. 1.5e-07;
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Yy 8 GCAGTTTCGCCCGGGGGGGGTGCAAGACGTTTCGGCCGACGGAGCGGAGCGGCTACGCGGG 67
```



```
FT      /cag- d
FT      /note- "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      /cag- e
FT      /note- "encodes a-y-l transferase (ATY) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /cag- f
FT      /note- "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /cag- g
FT      /note- "encodes a-y-l carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /cag- h
FT      /label- epoc_gene
FT      /note- "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /cag- i
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /cag- j
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /cag- k
FT      /note- "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /cag- l
FT      /note- "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /cag- m
FT      /note- "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /cag- n
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /cag- o
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /cag- p
FT      /note- "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /cag- q
FT      /note- "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /cag- r
FT      /note- "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /cag- s
FT      /note- "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /cag- t
FT      /note- "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /cag- u
FT      /note- "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /cag- v
FT      /note- "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /cag- w
FT      /note- "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /cag- x
```

```
FT      /label- epoc_gene
FT      /note- "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /cag- y
FT      /note- "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /cag- z
FT      /note- "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /cag- aa
FT      /note- "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /cag- ab
FT      /note- "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /cag- ac
FT      /note- "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /cag- ad
FT      /label- epod_gene
FT      /note- "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /cag- ae
FT      /note- "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /cag- af
FT      /note- "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /cag- ag
FT      /note- "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /cag- ah
FT      /note- "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /cag- ai
FT      /note- "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /cag- aj
FT      /note- "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /cag- ak
FT      /note- "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /cag- al
FT      /note- "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /cag- am
FT      /note- "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /cag- an
FT      /note- "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /cag- ao
FT      /note- "encodes DH5"
FT      misc_RNA
FT      29433..30287
FT      /cag- ap
FT      /note- "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /cag- aq
FT      /note- "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /cag- ar
FT      /note- "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /cag- as
FT      /note- "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /cag- at
FT      /note- "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /cag- au
```

FT		/note= "encodes DH6"	
FT	misc_RNA	35760..36641	
FT		/*tag= av	
FT		/note= "encodes ER6"	
FT	misc_RNA	36705..37256	
FT		/*tag= aw	
FT		/note= "encodes KR6"	
FT	misc_RNA	37470..37769	
FT		/*tag= ax	
FT		/note= "encodes ACP6"	
FT	CDS	37912..49308	
FT		/*tag= ay	
FT		/label= epoE_gene	
FT		/note= "encodes modules 7 and 8"	
FT	misc_RNA	38014..39375	
FT		/*tag= az	
FT		/note= "encodes KS7"	
FT	misc_RNA	39589..40626	
FT		/*tag= ba	
FT		/note= "encodes AT7"	
FT	misc_RNA	41341..41922	
FT		/*tag= bb	
FT		/note= "encodes KR7"	
FT	misc_RNA	42181..42423	
Query Match 100.0%; Score 101; DB 21; Length 71989;			
Best Local Similarity 100.0%; Pred. No. 9e-16;			
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	1	GATGACAGCGCTTTGTGCCCGGGCGGCGTGCACAACGCTTCGTGCCCGACGCGAGCGCTA	60
Db	21350	GATGACAGCGCTTTGTGCCCGGGCGGCGTGCACAACGCTTCGTGCCCGACGCGAGCGCTA	21409
OY	61	CGCGCGGCGCGAGGCGTGCGCCGTGTGTGCTCAAGGCGC	101
Db	21410	CGCGCGGCGCGAGGCGTGCGCCGTGTGTGCTCAAGGCGC	21450
RESULT 2			
ID	AAZ55887		
XX	AAZ55887 standard; DNA; 68750 BP.		
AC	AAZ55887;		
DT	10-APR-2000 (first entry)		
XX	Sorangium cellulosum 68.75 kb contig.		
DE			
XX	Epothilone biosynthesis; type I polyketide synthase; taxol substitute; anticancer; ds.		
KW			
RW			
XX	Sorangium cellulosum.		
OS			
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1826	
FT		/*tag= a	
FT		/partial	
FT	CDS	/product= "Partial Orf 1 protein (AAZ58580)"	
FT		/note= "No initiation codon given in the specification"	
FT		complement (1900..3171)	
FT		/*tag= b	
FT	CDS	/product= "Orf 2 protein (AAZ58581)"	
FT		3415..5556	
FT		/*tag= c	
FT	CDS	/product= "Orf 3 protein (AAZ58582)"	
FT		complement (5612..5592)	
FT	CDS	/*tag= d	
FT		/product= "Orf 4 protein (AAZ58583)"	
FT		6226..6675	
FT		/*tag= e	
FT	CDS	/product= "Orf 5 protein (AAZ58584)"	
FT		7610..11875	
FT	CDS	/*tag= f	

FT	/product= "Type I polyketide synthase, EPOS A
ET	(AAV58573)"
FT	11872..116104
FT	/*tag= g
FT	/product= "Non-ribosomal peptide synthetase, EPOS P
FT	16251..21749
FT	/*tag= h
FT	/product= "Type I polyketide synthase, EPOS B
FT	(AAV58575)"
FT	21746..43519
FT	/*tag= i
FT	/product= "Type I polyketide synthase, EPOS C
FT	(AAV58576)"
FT	43524..54920
FT	/*tag= j
FT	/product= "Type I polyketide synthase, EPOS D
FT	(AAV58577)"
FT	54935..62254
FT	/*tag= k
FT	/product= "Type I polyketide synthase, EPOS E
FT	(AAV58578)"
FT	62369..63628
FT	/*tag= l
FT	/product= "Cytochrome P450 oxygenase homologue, EPOS F
FT	(AAV58579)"
FT	63779..64333
FT	/*tag= m
FT	/product= "Orf 6 protein (AAV58585)"
FT	complement (63853..64290)
FT	/*tag= n
FT	/product= "Orf 7 protein (AAV58586)"
FT	64363..64920
FT	/*tag= o
FT	/product= "Orf 8 protein (AAV58587)"
FT	complement (64287..64727)
FT	/*tag= p
FT	/product= "Orf 9 protein (AAV58588)"
FT	65063..65767
FT	/*tag= q
FT	/product= "Orf 10 protein"
FT	complement (65008..65874)
FT	/*tag= r
FT	/product= "Orf 11 protein (AAV58590)"
FT	complement (65871..66338)
FT	/*tag= s
FT	/product= "Orf 12 protein (AAV58591)"
FT	66667..67137
FT	/*tag= t
FT	/product= "Orf 13 protein (AAV58592)"
FT	67334..68251
FT	/*tag= u
FT	/product= "Orf 14 protein (AAV58593)"
FT	68346..68750
FT	/partial
FT	/*tag= v
FT	/product= "Partial Orf 15 protein (AAV58594)"
FT	/note= "No termination codon given in the specification"
XX	
PX	W09966028-A2.
PX	
PD	23-DEC-1999.
PX	
PE	16-JUN-1999; 99WO-EP04171.
XX	
PR	18-JUN-1998; 98US-0099504.
PR	24-SEP-1998; 98US-0101631.
PR	05-FEB-1999; 99US-0118906.
XX	
PA	(NOVS) NOVARTIS AG.
RA	(NOVS) NOVARTIS-ERFINDUNGEN VERM GBS MBH.
P1	Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX	

XX WPI: 2000-097741/08.
DR DR P-PSDB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58589, AAY58591,
XX AAY58592, AAY58593, AAY58594.
PT New isolated epoethione synthase genes, used for the recombinant
PT production of epoethione for use in cancer therapy
PS Claim 14; Page 87-104; 174pp; English.
XX This sequence represents a 68.75 kb contig from *Sorangium cellulosum*
CC comprising 22 open reading frames (ORFs) and includes genes encoding
CC proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPDS A (AAY58573) and EPDS P (AAY58574) are involved in formation of
CC the thiazole ring formation of epoethiones, and EPDS B, EPDS C, EPDS D
CC and EPDS E (AAY58575-58578) are involved in polyketide backbone
CC formation. EPDS F (AAY58579) is an epoethione macrolactone oxidase, and
CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epoethiones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epoethiones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epoethiones as anticancer agents, they are problematical to produce on a
CC large scale. Epoethiones are too complex for industrial scale chemical
CC synthesis, and *Sorangium cellulosum* is difficult to ferment, producing
CC poor yields of epoethiones. The nucleic acids of the invention may be
CC used for the recombinant production of epoethiones in a heterologous host
CC that is more amenable to fermentation.
XX
SQ Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
Query Match 98.4%; Score 99.4; DB 21; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.2e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GATGACGCGCTTCGCCGCGGCGGCTGCAGACGTTCTCGCCGACGCGACGCGTA 60
DB 26962 GATGACGCGCACTTCGCCGCGGCGGCTGCAGACGTTCTCGCCGACGCGACGCGTA 27021
QY 61 CGCGCGGCGCGAGGCTGCGCGCTGCTGCTCAAGCGGC 101
DB 27022 CGCGCGGCGCGAGGCTGCGCGCTGCTGCTCAAGCGGC 27062
RESULT 3
AAI99683/c
ID AAI99683 standard; DNA: 4403765 BP.
AC AAI99683;
XX 15-JAN-2002 (first entry)
DE Mycobacterium tuberculosis strain H37Rv genome SPO ID NO 2.
XX
KM Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 9805-0103840.
XX

XX
PR 24-JUN-1998; 9805-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
DR WPI: 2001-647261/74.
XX
PT Evaluating strain variation of *Mycobacterium tuberculosis*, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where *M. tuberculosis* strains CDC
PS 1551 and H37Rv differ
XX
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC *Mycobacterium tuberculosis* or related *Mycobacterium* by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docid=6294328b1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
Query Match 62.0%; Score 62.6; DB 22; Length 4403765;
Best Local Similarity 79.6%; Pred. No. 8.9e-07;
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 8 GCGCTTCGCCGCGGCGGCTGCAGACGTTCTCGCCGACGCGGCTGACGCGCG 67
DB 4252406 GTCTGCGCGCGGCGGCGGCTGCAGACGTTCTCGCCGACGCGGCTGACGCGCG 4252347
QY 68 GCCGAGGCGTCCGCGCTGCTGCTCAAGCGG 100
DB 4252346 TCCGAGGCGCGGCGGCTGCTGCTCAAGCGG 4252314
RESULT 4
AAI99682/c
ID AAI99682 standard; DNA: 4411529 BP.
AC AAI99682;
XX 15-JAN-2002 (first entry)
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KM Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 9805-0103840.
XX
PR 24-JUN-1998; 9805-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
DR WPI: 2001-647261/74.
XX

XX Evaluating strain variation of *Mycobacterium tuberculosis*, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where *M. tuberculosis* strains CDC
PT 1551 and H37Rv differ.

XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC *Mycobacterium tuberculosis* or related *Mycobacterium* by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other:

Query Match 62.0%; Score 62.6; DB 22; Length 4411529;
Best Local Similarity 79.6%; Pred. No. 8.9e-07;
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 8 GCGCTTGGCCCGGCGGCGGTGCAAGAGCTTCTGCGCCGACGCGAGCGCTACGCGCG 67
DB 4260150 GTGCTGCGCGCCGACGCGCGGATCAAGCTTCACGCCGACGCGCTACACCGCG 4260091
OY 68 GCGGAGGCGTGGCGGTGGTGTGCTCAAGCGG 100
DB 4260090 TCCGAGGCGCGGCGATGCTGCTCAAGCGG 4260058

RESULT 5
ID AAS17367 standard; DNA: 33529 BP.
XX AAS17367;
XX
DT 12-MAR-2002 (first entry)
XX
DE DNA sequence of *S. cellulosum* polyketide synthase cosmid, PKOS28-26.
XX
KW Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW acyl transferase domain; dehydratase domain; ketoreductase domain;
KW acyl carrier protein domain; PKOS28-26; ds.
XX
OS *Sorangium cellulosum*.
XX
PN US6280999-B1.
XX
PD 28-AUG-2001.
XX
PE 31-AUG-1998; 98US-0144085.
XX
PR 22-JAN-1998; 98US-0010809.
XX
PA (KOSA-) KOSAN BIOSCIENCE.
XX
PI Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX
DR WPI; 2001-606536/69.
XX
PT Novel purified, isolated DNA molecule from *Sorangium cellulosum* having
PT polyketide open reading frame encoding modules with one or more domains
PT such as ketosynthase, acyl transferase and acyl carrier protein domains
PT
XX
PS Claim 4; Fig 1; 72pp; English.

XX
CC The present invention relates to the isolation of novel *Sorangium*
CC *cellulosum* polyketide synthases (PKS), and the polynucleotide sequences
CC encoding them. The polyketide synthases include catalytic domains such
CC as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC ketoreductase domain and acyl carrier protein domain. A host cell
CC comprising a PKS ORF (open reading frame) which encodes one or more
CC more PKS domains is useful for producing polyketide synthases from which
CC polyketides can be produced. The host cells are useful for constructing
CC a library, where each individual colony of the library represents a
CC colony with the ability to produce a particular PKS synthase and
CC ultimately a particular polyketide. The polyketides produced by these
CC colonies can be used collectively in a panel to represent a library or
CC may be assessed individually for activity. Colonies in the library are
CC also induced to produce the relevant synthases and thus to produce the
CC relevant polyketides to obtain a library of candidate polyketides which
CC can be screened for binding to desired targets such as receptors,
CC signalling proteins, etc. The present sequence represents the DNA
CC sequence of cosmid PKOS28-26 which encodes one or more domains of
CC *S. cellulosum* PKS.
CC Note: The present sequence is said to encode the functional domains
CC of *S. cellulosum* PKS which correspond to domains or domain subsets of
CC the amino acid sequences of ORF1 (AA010700) and ORF2 (AA010701).

XX
SQ Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other:

Query Match 61.0%; Score 61.6; DB 23; Length 33529;
Best Local Similarity 76.0%; Pred. No. 2.2e-06;
Matches 76; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 2 ATGCACGCGCTTGGCCCGGCGGCGGTGCAAGAGCTTCTGCGCCGACGCGGAGCGCTAC 61
DB 10349 ACGCTGCGGTCTCGCCGACGAGCGGTGCAAGAGCTTCTCGCAGCGCGGCGAGCTAC 10408
OY 62 GCGCGGCGCGAGGCGTGGCGCGCTGCTGCTCAAGCGCG 101
DB 10409 GCGCGGCGCGAGGCGCGTGGGTGGTGTGATGCGTC 10448

RESULT 6
ID ABO62819 standard; DNA: 333 BP.
XX ABO62819;
XX
DT 16-AUG-2002 (first entry)
XX
DE *Mycobacterium tuberculosis* BAC vector clone RV272r7.
XX
KW *Mycobacterium tuberculosis*; *Mycobacterium bovis*; *Mycobacterium*;
KW detection; BAC vector; bacterial artificial chromosome; tuberculosis;
KW gene; ds.
XX
OS *Mycobacterium tuberculosis*.
XX
PN WO9954487-A2.
XX
PD 28-OCT-1999.
XX
PE 16-APR-1999; 99WO-1B00740.
XX
PR 16-APR-1998; 98US-0060756.
XX
PA (INSP) INST PASTEUR.
XX
PI Cole S, Buchrieser-Brosch R, Gordon S, Billault A;
XX
DR WPI; 2000-013262/01.
XX
PT Isolation of polynucleotides from *Mycobacterium* genomes, useful for
PT detection of *Mycobacterium* and for combating tuberculosis -
PT
XX
PS Claim 23; Page 77; 161pp; English.

CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
CC
SQ Sequence 5485 BP; 896 A; 1745 C; 1926 G; 918 T; 0 other;
Query Match 58.8%; Score 59.4; DB 22; Length 5485;
Best Local Similarity 77.4%; Pred. No. 8.4e-06;
Matches 72; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 8 GCGCTTTCGCCCGCGCGGTGCAAGAGCTTCTCGCGCGACGCGGCTACGCGG 67
DB 751 GCGCTGTACCGACCGCGCGCGATGCAACAGCTTCGACGCCACGCGGATGCTCGCGC 810
QY 68 GCCGAGGCGCTGCCGCGGTGCTGCTCAAGCGG 100
DB 811 GCGGAGGCGCTGCCGCGGTGCTGCTCAAGCGG 843
RESULT 9
AAF89970
ID AAF89970 standard; DNA: 758 BP.
AC AAF89970;
XX
DT 06-AUG-2001 (first entry)
XX
DE Partial nucleotide sequence of a type I polyketide synthase.
XX
KM Metabolic pathway operon; polyketide; polyketide antibiotic;
XX type I polyketide synthase; ss.
OS Unidentified.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PE 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanne P, Pernodet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francon F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
DR P-PSDB; AAB83962.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Claim 35; Page 222; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange

CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC encode partial type I polyketide synthases, and were isolated using the
CC method of the invention.
CC
SQ Sequence 758 BP; 121 A; 250 C; 247 G; 140 T; 0 other;
Query Match 56.4%; Score 57; DB 22; Length 758;
Best Local Similarity 74.2%; Pred. No. 3.6e-05;
Matches 72; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 5 CACGGCTTTCGCCCGCGCGGTGCAAGAGCTTCTCGCGCGACGCGGCTACGCG 64
DB 469 GCGATGTGGCCCGCGGAGCGAGCTGCAAGAGCTTCGACGCCGCGCGGATGATG 528
QY 65 CGGCGGAGGCGCTGCCGCGGTGCTGCTCAAGCGG 101
DB 529 CGCGGAGGCGCTGCCGCGATCTGCTGCTGCAAGCGG 565
RESULT 10
AAF90038
ID AAF90038 standard; DNA: 5088 BP.
AC AAF90038;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
KM Metabolic pathway operon; polyketide; polyketide antibiotic;
XX type I polyketide synthase; ss.
OS Unidentified.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PE 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeanne P, Pernodet J, Guerin M, Simonet P, Courtois S;
PI Cappellano C, Francon F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
DR P-PSDB; AAB83975.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Claim 35; Page 309-311; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;

OY 70 CGAGGCTGCGCGGCTGCTCTCAAGCGC 101
 DB 822 CGAGGCTGCGCGGCTGCTCTCAAGCGC 853

Search completed: November 5, 2002, 13:04:56
 Job time : 1762.22 secs

RESULT 15
 ABO63181/C
 ID ABO63181 standard; DNA: 507 BP.

XX ABO63181;

XX 16-AUG-2002 (first entry)

DE Mycobacterium tuberculosis BAC vector clone RV907.

XX Mycobacterium tuberculosis; Mycobacterium bovis; Mycobacterium;

KM detection; BAC vector; bacterial artificial chromosome; tuberculosis;

XX gene, ds.

OS Mycobacterium tuberculosis.

XX WO954487-A2.

PD 28-OCT-1999.

XX 16-APR-1999: 99MO-IB00740.

PR 16-APR-1998: 98US-0060756.

XX (INSP) INSP PASTEUR.

PI Cole S, Buchrieser-Brosch R, Gordon S, Billault A;

XX MPI: 2000-013262/01.

PT Isolation of polynucleotides from mycobacterial genomes, useful for

PS detection of Mycobacteria and for combating tuberculosis.

XX Claim 23: Page 121; 161pp; English.

CC The present invention describes a method for isolating a polynucleotide
 CC of interest that is present or is expressed in a genome of a first
 CC mycobacterium strain and that is absent or altered in a genome of a
 CC second mycobacterium strain, which is different from the first strain
 CC using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
 CC vectors, which are preferably immunolysed, can be used to detect
 CC mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
 CC samples. The polynucleotides identified are useful as probes or primers
 CC for detecting a given mycobacterium of interest. By aligning the
 CC polynucleotides contained in the recombinant BAC vectors it is possible
 CC to physically map a polynucleotide of mycobacterial origin in a
 CC biological sample. The methods and vectors from the present invention
 CC are useful in providing information for combating tuberculosis. It is
 CC possible to compare genomes between different strains or species and
 CC their non-pathogenic strains or species counterparts. ABO62497 to
 CC ABO63228 and ABB81227 to ABB81230 represent sequences used in the
 CC exemplification of the present invention.

XX Sequence 507 BP; 89 A; 176 C; 132 G; 87 T; 23 other:

Query Match 55.8%; Score 56.4; DB 21: Length 507;
 Best Local Similarity 74.2%; Pred. No. 5.2e-05;
 Matches 69; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 8 GCGCTTTCGCGCGCGGCTGCTCAAGAGCTTCGCGCGCGAGCGGCTACGCGCG 67
 DB 204 GCGCTGTCACCGACGACGCGATGTCNACAGTTTGACGCCAACCGGATGATTCGTGCGC 145

OY 68 GCGAGGCGCGCGCGGCTGCTCAAGCGC 100
 DB 144 GCGAGGCGCGCGGCTGCTCAAGCGC 112

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 458.785 Seconds

(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_21350_21450

Perfect score: 101
Sequence: 1 gatgcacgcgcttcgcgcg.....cgtgtgtgtgtcctaagcgcgc 101Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pla:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	58733	1	AF217189 Sorangium
2	101	100.0	71989	6	AF172664 Sequence
3	99.4	98.4	68750	1	AF210843 Sorangium
4	99.4	98.4	68750	6	AR193029 Sequence
5	99.4	98.4	68750	6	AR199551 Sequence
6	99.4	98.4	68750	6	AR199559 Sequence
7	99.4	98.4	68750	6	AR199567 Sequence
8	99.4	98.4	68750	6	AR201097 Sequence
9	99.4	98.4	68750	6	AR208671 Sequence
10	80.8	80.0	10910	1	AX024383 Sequence
11	80.8	80.0	10910	6	AX024276 Sequence
12	64.2	63.6	27522	1	AB070942 Streptomy
13	63.8	63.2	66808	1	SA0421825 Stigmatel
14	62.6	62.0	14634	1	AE007184 Mycobacte
15	62.6	62.0	23740	1	MTV026 Mycobacte
16	62.2	61.6	42603	1	AF188287 Stigmatel
17	61.6	61.0	33529	6	AR166425 Sequence
18	61.2	60.6	82746	1	AF453501 Actinosyn
19	59.6	59.0	30000	6	AX250263 Sequence
20	59.4	58.8	333	6	AR128456 Sequence
21	59.4	58.8	5481	6	AX069075 Sequence
22	59.4	58.8	20679	1	MTCY19H9 Mycobacteri
23	59.4	58.8	32870	1	AE007122 Mycobacte
24	59.2	58.6	19601	1	AB070944 Streptomy
25	58.8	58.2	49736	1	AF319998 Stigmatel
26	58.4	57.8	213050	1	AL646079 Ralstonia
27	58.2	57.6	90348	1	AF497482 Micromona
28	57	56.4	758	6	AX153712 Sequence
29	57	56.4	5088	6	AX153795 Sequence
30	57	56.4	34071	6	AX153790 Sequence
31	57	56.4	42717	6	AX153789 Sequence
32	56.8	56.2	764	6	AX153716 Sequence
33	56.8	56.2	6564	6	AX069073 Sequence
34	56.4	55.8	507	6	AR128818 Sequence
35	56.4	55.8	3185	1	AF41157364 Actinomad
36	56.4	55.8	3900	1	AF262754 Amycolato
37	56.4	55.8	43280	1	SF078289 Streptomyce
38	56.4	55.8	78210	1	AB070949 Streptomy
39	55.6	55.0	3573	1	AB016763 Streptomy
40	55.2	54.7	671	6	AX153719 Sequence
41	55.2	54.7	189050	1	AL646066 Ralstonia
42	54.8	54.3	53784	1	AMW223012 Amycolato
43	54.8	54.3	53789	6	AE9720 Sequence 3
44	54.8	54.3	90445	1	AF040570 Amycolato
45	54.6	54.1	11219	1	SERERYAA Saccharopol

ALIGNMENTS

```
RESULT 1
AF217189
LOCUS
DEFINITION AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION AF217189
VERSION AF217189
KEYWORDS
SOURCE
ORGANISM Polyangium cellulosum.
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)
```

Pred. No. is the number of results predicted by chance to have a


```
QSTRIMETTLDLKRLPWLGDHVGAVVFPGAATLEMATSSGAELDGPLOTIDV
LAELAFAGDAVLVOYVTTBOPSGRLOFOLASRPAQOAGHSEFVHARGALLVRTE
VPAGITTSARVARLOASMPAATYATLEMTLOQRPAGOGHSELRGSGALGRVRLE
DAGSAATYRIHPALDLDCFOYVGSIFRAGGEBATWVVEGSLDLRPSGELCHRA
RVNHGRQTPROGADENWVDSSGAVVAEVLVAQRLPVGRRREEDDWLEETLEMER
AAVETAKVNAWRWLLGGGGGIGALRSMLEAGVLAHAESNT SAAGVALLAKAF
DGOAPTAVHLGSLDGGELDPGLGAGALAPRADVSPALDPLALHRRVOCVLPV
OALMGFRDAPRLMLTRGAQAVAGVSVTOAPLGLGRTVIMEHADLCARGDLID
PTRPDGELGALLAELADDAFEVALRGECVARIYRQETRRRGIESCVPTDYI
IRADSTIYVGTGGLGSLVAGMLAEKAGHDVLVGRSAAASVEGRVAALAEKAR
VTVAKADVADRAQLEIRLEVTTSCMPTRGVHAAGILDDGLMORTYIRARKVAPK
VOGALHIALTRAPLSEFFVLVYASGVGLSGPGGONAAATFLIALHHRRAOCLPA
LSVMDGLEAEVMAAABEDRGARLVSRGMSRTTDEGSAALRIEESRVGQVLEMT
PRIMEVTPAASSRMLSLRYLTHASAGAGAGOGALDRLRLAAHRSARSGILEPILR
AOISOVLREPKGTFVDPARLTSLGKNSLMGELREIRITAMGITVPAULMTPTVVA
LSGHLARACCAAPVESPHTTADSAAVELEMSODDLTOLIAKERALT"
gene
16134..37907
/contig="epod"
/contig_start=1
/contig_end=16134
/contig_size=16134
/contig_type="I polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQNPLKOAIIITOREERLAGLAQALETERPIAIV
GIGRFGGADAPFAFWELLDAERDAVPLDRMALVGVAPVAPVHAGLITEPIDC
FDAFFGISPRESAIDPHRLLEVAEGEDGIPRSIDGATGVFAVATFADVA
RTVARLPREEDAYSATGNMLSTIAGRSITGLPGCTVATDASSLVIAHACRS
LRAGSDALAGVSTLSPDMMEARTOALSPGRCRTDASNGVREGGCLIV
LKRLSDAQRDGRIMALIRGSAINHDSSTGCTANVTAQETVLRREALRSAREGAV
DYVETHTGISTGDETEVALVRAVGPARGSGRCVLAATNIGHLPAAGVAGLILK
AALSLTHERIPRNLNRTLNRIIRLEGSALALATEPVPRTDRFRFAGVSGFMSGT
NAHVLEAPVAPVLEPAPERSAELVLVSGEGALDQAARLEHLMHPELIGDY
AFSLATTSAMTHRLAVATSRREGILALISVAGOGTAPGAARLJASSRGELALFT
GOGAOTPGMGRIICAMPAPREAPRCVTLDRDELDRALREVMASAFSLLIADRG
APTOPALFAVEYALTAIMRSGVPELIVGHSIGLVAACGVTSEDDGVRVLAARG
RLMGSLSGAMVSLGAPAEVAAVAPHAAMVSAVAVNGEQVIAVEQVAVQVIAIA
GFAAGVTRKLVSHAFSHPLMEPEMEEFGRVAVSVTRYRPSVSLVNSLKGVTDE
Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 3e-10;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGCACGGCGCTTTCGCCGGGGGGGCGTCAAGAGCTTCTCGCCGACGGACGGCTA 60
|||||
DB 21350 GATGCACGGCGCTTTCGCCGGGGGGGCGTCAAGAGCTTCTCGCCGACGGACGGCTA 21409
QY 61 CGCGCGGGCGGAGGGCTGCGCGTGGTCTCAAGCGGC 101
|||||
DB 21410 CGCGCGGGCGGAGGGCTGCGCGTGGTCTCAAGCGGC 21450
RESULT 2
AR172664
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342 A 2 16-OCT-2001;
FEATURES
source 1..71989
Location/Qualifiers
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
```

```
ORIGIN
Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.8e-10;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGCACGGCGCTTTCGCCGGGGGGGCGTCAAGAGCTTCTCGCCGACGGACGGCTA 60
|||||
DB 21350 GATGCACGGCGCTTTCGCCGGGGGGGCGTCAAGAGCTTCTCGCCGACGGACGGCTA 21409
QY 61 CGCGCGGGCGGAGGGCTGCGCGTGGTCTCAAGCGGC 101
|||||
DB 21410 CGCGCGGGCGGAGGGCTGCGCGTGGTCTCAAGCGGC 21450
RESULT 3
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain so ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Mycococcales; Sorangiineae; Polyangiaceae; Polyangium.
1 (bases 1 to 68750)
AUTHORS Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milmanow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Struelens,A., Cyr,D.D.,
Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Iigou,J.M.
TITLE The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
Chem. Biol. 7 (2), 97-109 (2000)
JOURNAL Molnar,I.
MEDLINE Submitted (03-DEC-1999) Natural Product Genetics, Novartis
PUBMED Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box
12257, Research Triangle Park, NC 27709, USA
REFERENCE
JOURNAL
TITLE
AUTHORS
JOURNAL
FEATURES
source 1..68750
Location/Qualifiers
1. 68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/contig="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASIDALFARATSAVLDGSHGATERRHYLAERGITEDLRALREH
LRIOGSPSPHCMTGDLTVELADNDOLASTSPHNASISHPDPTSNAMVDPGATV
RWLARGAPLURETEEBERERERKQERRLMLAARPCFADLREEDDANGDLGP
MSPEVAEERRLRASVATPELACALALMLTGAGPMSGVAIYELPILNGLRGLPT
AIAASAGPTESEALRGARLFAESWEVVSRSQGNIPDALMERLRTIVAMGADN
LSRPERAIAAEVRLRAOPAPFAAGAGLAVAGSSGRLGLVTDGDAYSGDND
ITMRQGRISPVVILAGNDPPELAPLPSOMLVANAHAGTISKVLTGSPLIYARN
OARPSLYHARGFANVQNAVPPDERGAPVYVQSTIMEFEHPPTRIHPAGSAPS
LACDEHLIWCLELSAGRLMLWHPHHRGAPSRFAYLGEHPPLATWYSLTNATHIL
WADPDRALILGDKRTGVEPIVLAETRHPAHVYSEDRIALFGQDPSRDMWHEHIL
SGASTVADYOROLDPRDPMVILNRGLETFTNDRILTLARS"
complement(1900..3171)
/contig="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to Dd-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
```


Best Local Similarity 99.0%; Pred. No. 5.6e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 60
Db 26962 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 27021

Qy 61 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 27062

RESULT 4
LOCUS ARI93029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.6e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 60
Db 26962 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 27021

Qy 61 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 27062

RESULT 5
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.6e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 60
Db 27022 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 27062

Db 26962 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 27021

Qy 61 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 27062

RESULT 6
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 635458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 635458-A 1 12-MAR-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.6e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 60
Db 26962 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 27021

Qy 61 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 27062

RESULT 7
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.6e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 60
Db 26962 GATGACGCGCTTTCGCCCGGGCGGCTGCAAGACGTTCTCGGCCGACGGAGCGGCTA 27021

Qy 61 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 101
Db 27022 CGCGGGGCGCAGAGGCTGCGCGGTGTGTGCTCAAGCGGC 27062

RESULT 8
LOCUS AR201097 68730 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison , Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25339 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.6e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGACGGCGCTTTCGCCGGCGGGGCGGTCGACAGAGCTTCTCGCGCGAGCGGAGGCTTA 60
|||||
Db 26962 GATGACGCGACTTTCGCCGGCGGGGCGGTCGACAGAGCTTCTCGCGCGAGCGGAGGCTTA 27021
|||||

QY 61 CGCGGGGGCGGAGGCGTGCCTGTGTGTCTCAAGCGGC 101
|||||
Db 27022 CGCGGGGGCGGAGGCGTGCCTGTGTGTCTCAAGCGGC 27062
|||||

RESULT 9
LOCUS AR208671 68753 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6183787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethliones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25339 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5.6e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGACGGCGCTTTCGCCGGCGGGGCGGTCGACAGAGCTTCTCGCGCGAGCGGAGGCTTA 60
|||||
Db 26962 GATGACGCGACTTTCGCCGGCGGGGCGGTCGACAGAGCTTCTCGCGCGAGCGGAGGCTTA 27021
|||||

QY 61 CGCGGGGGCGGAGGCGTGCCTGTGTGTCTCAAGCGGC 101
|||||
Db 27022 CGCGGGGGCGGAGGCGTGCCTGTGTGTCTCAAGCGGC 27062
|||||

RESULT 10
LOCUS AX024383 10910 bp DNA linear BCT 15-SEP-2000
ORIGIN

DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024383
VERSION AX024383.1 GI:10184587
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 10910)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source Location/Qualifiers
1..10910
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 1852 a 3124 c 3613 g 2321 t
ORIGIN

Query Match 80.0%; Score 80.8; DB 1; Length 10910;
Best Local Similarity 88.0%; Pred. No. 2.5e-06;
Matches 88; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 ATGCAGCGCTTTCGCCGGCGGGGCGGTCGACAGAGCTTCTCGCGCGAGCGGAGGCTTA 61
|||||
Db 2895 ATGCAGCTTTCGCCGCAAGATGGCGGTGCACAGAGCTTCTCGCGCGAGCGGAGGCTTA 2954
|||||

QY 62 GCGGGGCGGAGGCGTGCCTGTGTGTCTCAAGCGGC 101
|||||
Db 2955 GCGGGGCGGAGGCGTGCCTGTGTGTCTCAAGCGGC 2994
|||||

RESULT 12
LOCUS AB070942 27522 bp DNA linear BCT 11-OCT-2001
DEFINITION Streptomyces avermitilis polyketide-2 biosynthetic gene cluster.
ACCESSION AB070942
VERSION AB070942.1 GI:15823997


```

gene
  RGTADDCDALQRFDELVAGATTAEDQT"
  1980..3314
  /gene="RV3798"
  /note="RV3798, (MT026.03), len: 444. Putative
  transposase identical to IS1557 MTCY373_32,
  sp1010621.Y02X_MYCVU_HYPOTHEICAL_50.2_KD_PROTEIN_CV37
  (444 aa), 99.8% identity in 444aa overlap and similar to
  bacterium (163 aa), FASTA scores: opt: 767 z-score: 926.8
  E(1: 0, 67.3% identity in 168 aa overlap, or to B151001_1
  B. paraperlensis insertion sequenc (406 aa), 24.9% identity
  in 402 aa overlap."
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein RV3798"
  /protein_id="CA17862.1"
  /db_xref="GI:2950417"
  /translation="MKRVILFRALIGDKRTVIEDIEFEEDADGCAVYIARVPSA
  VLRGCRGKRWSDRG;GLRQWRSIDMGTYEVFLAEAPRNCPTHPPTVAVPWA
  RHAGHTYAFDDTAVWLVA;ACSKTAVCELMIRMTVGAIVARVADTEKRIDREANL
  RRIQDEISIKRHHRYLT;VVDHSGRLVMAAPGHDKATLGLEFDDAGRAHQITHV
  SADADMIADVYTERCPD;IQCADPFFVMAATEALDVERRRAMNDARAIARPEPVG
  RGRPGKNAAPRPERERAR;LKGARVALMKNPEDLTERQAKLIMAKTDPRLYRAVL
  KESLRHVFSGKEGKQALDWMISMAORCRIPVEVFLARIKRHRVAIDALDHGLSQ
  GLIESNTKIRILRTIAFIFRSQALIALAMLTLAGHRLPLGRNHNHPLDSQ"
  1980..3311
  /gene="RV3798"
  /note="IS1557-3rd copy"
  complement(3382..4935)
  /gene="accD4"
  complement(<382..4935)
  /note="RV3799c, (MTV026.04c), accD4, len: 517. Probable
  Propionyl-CoA carboxylase: Beta chain, similar to many,
  e.g. sp1P53003.IPCCE_SACER_PROPIONYL-COA CARBOXYLASE BETA
  CH (546aa), FASTA scores: opt: 1549 z-score: 1841.5 E(1:
  0; 50.7% identity in 519 aa overlap, Also similar to
  U00012_17 M. leprae cosmid B1308 (549 aa), 49.8% identity
  in 514 aa overlap and to MTCY71_20 M. tuberculosis cosmid
  (548 aa), 48.9% identity in 515 aa overlap"
  /codon_start=1
  /transl_table=11
  /product="accD4"
  /protein_id="CA17863.1"
  /db_xref="GI:2950418"
  /db_xref="SPTREMBL:053578"
  /translation="MTVTEPLVHTTAETAELEKLELAKKEGGEKAARKDKKGPS
  ARARITELVDPGSFMEIGALCRTPGDPNALYGGVVTGGLINGRFGVFSHQVFG
  GTVGEHFGKRVANLMECMVGCPIVIGINDSGARVQDPTSLAAVEAGRSRHEILSG
  LVPOISIIIGKCGAGVVEPIQDTDLVAVDQGYMVFDPDKVDTGDSVDELDELG
  ADHQASVGNHGVVESEAAVYVYVDFLSFPCNCDKPPVAVPGDPEPTGSDLELD
  STYPDNDNMAVDHVEVLRITFDGDFDLVAAQGGALITIGYAVDSDRTGVVANOVMH
  MSGAIDNEASDAARFIRSDAEDPLVAVGSKOLADINENPRTARINAGDCAQOLMKRFP
  VEAADVPKYITTRKSYGCAVAVGSKOLADINENPRTARINAGDCAQOLMKRFP
  DRAAPQAQIRKSFVENYHNLMAIPWIAERGTIDAVIDPHETLLRLKSMHLRDKQ
  LMRVGRKHG"
  complement(4932..10133)
  /gene="pkS13"
  complement(4932..10133)
  /gene="pkS13"
  /note="RV3800c, (MTV026.05c), len: 1733. Probable
  polyketide synthase, similar to many, e.g.
  gp1M636761SERERYAA.1 Sterytrae first ORF of eryA gene,
  involved in complex polyketide formation in erythromycin
  biosynthesis. FASTA scores: opt: 1329 z-score: 1411.0 E(1:
  0; 33.6% identity in 1232 aa overlap or to AF040570_9
  Amycolatopsis mediterranei rifamycin biosynthetic gene
  cluster (3413 aa), 34.3% identity in 1162 aa overlap.
  ALSO highly similar to mycobacterial putative polyketide
  synthase: MTCY338_19, U00023_3 M. leprae,

```

```

MTCY22610_2, MTCY19H9_2, U00023_6 M. leprae, contains
PS00606 Beta-ketoacyl synthases active site"
/codon_start=1
/transl_table=11
/product="pkS13"
/protein_id="CA17864.1"
/db_xref="GI:2950419"
/db_xref="SPTREMBL:053579"
/translation="MADVAESEOENAPAEELTPEMROWLNNYKAVGRAPSDIDE
SVPMFELGSSPDADYAMADIEDLIGYILSVAVAAHPTISLARITIEGREGSLAG
DDEDMSTRGPEREDIRATVIGSTRPGEMNPEPTQWOLLEGRRGITDLEGRSEF
LEPERLAARVAGARRGGYLKIDKIGPDSFEFVARTADNDIDQRMALDELWELH
AIPASSLNGQAVVYIGSSTDYSLFVAVSDTVAAHPAITGTSSTSIANRSTYDF
HGSVYTIIDPACSSSLVAIHQVQALRNGADVAVAGNALTTPWTGTFEIGAVLA
PDGRKSFSDADGYTRSECGMLVLKRVDDARQGDALIAVLAGAVNHGSRNGLI
APNODQADVLRRAKYRAGIDIPRTVYIEAHGTGTLIDPSINFAGSPDYIDPDMRLKI
PALIGAVKTNVGLHSAGCAAMAKVYVLAJOHDKLPSPINFAGSPDYIDPDMRLKI
TPTDMPRRYGVALAGVSGSGGANNAHVYREVLPRDVEPEPEPEPEPEAAAPAE
APTLAGHARPRPEGNITDSDAVAEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE
VPLAVSAFTLSRKKAAMAFELADMMQSPSEQASLSIGRSLSRHHGSRVAVLAHDH
DEALIGLRVAAKQAPNVEFSDGVVTPGVVWLVGFQGHKKKRSYLRNEFAAM
IKRVVALVQDELGYVLELIDDDADYGLIETTVQVTFALQALGELRHHGAKPAVLI
GQSLDEAASVAFAGLSLMDATRALCSHLMGEBAALFGEYIRLMAVLESADDIR
EYFSDPDLQVGVYAPDTQVYIGGPEQDAILARAEKGRKATKAGSHTSQMD
PLIGELTALQELIKPTSPGCFSTYHBEERYIKPGGEIHDVEYWKKGLRSVYTFG
IRNAVDSGHTTELELAPNVALMOVALTADGLHDAOILPTLAKODEVSMTMA
OLYVGHODIDRTLEFSRASGPDQVANIPTPRKRREHMLPAHFGSDGSTMGPVAL
PGRRHYVEAPRPGDGVNDLAIVRAAAHVLDPQOLTLAEDORAVPDGARRVYTPMRHP
GCAVQVHARRIDESPTLYYDALYSRAGSESVYPTVGAATAAVAVDCAVPVPERAD
ADRETLSDSLTTRIMPSCMTRNSPDSEGTIAERLILVGSAGTYREDUPMEVPLEL
GUDSLMAVIRKRRVEDPDPPIQLPVAVDALVAVLIEVAVHREVOULHEHOK
TQVAEIARAQAEELLHGKVKTEPVDSEGVALLPSPQNEQPNPGLPAAVNDVPRDA
ABRVTFATWAIYTGKSPGIFENELRDLDEAAKTAQRLSERAEQPIYAEVLISSNI
EALADKRVTYLQAGQIDGFVTRLRAPPEAGGVVPEVPHRAGSVTVYEPILGRLPAD
TPMGFERVEGSIIEERAOQYVPLIKEMOGDGVYLVGMSGLGVLAACAGIGRLRQD
VPEVGLIDAVRAGEEIPQTEIKRKRMRQYAFKATFNVTIPALPYBOLELDLDEGO
VREVIDAVSOSGVQIPAGITIEHQRSTSDNRAIDPAQIDPDGHVTLTMADRYHDAT
MEPEPRAVQPGCGEYVSDLEVPPIGEGHQAIDEPITAKVGHMSRALQIDADR
TSEVGK"
complement(4945..4948)

```

RBS

complement(4945..4948)

Query Match 62.0%; Score 62.6; DB 1; Length 23740;
 Best Local Similarity 79.6%; Pred. No. 0.0049;
 Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY	8	GGCGTTTCGGCGGGGGGGGTGCGAGAGCGTTCGCGGAGCGGAGCGGTACGCGGGG	67
DB	9140	GTGCTGGCGCCCGACGGCCGCGATCAAGCTGCTTCAGCGGAGCGGCGGCTACACCGG	9081
QY	68	GGCGAGGGCTGCGCGGTGGTGCTCAAGCGG	100
DB	9080	TCCGAGCGGCGGCGCATGCTGCTCAAGCGG	9048

Search completed: November 5, 2002, 12:16:41
 Job time : 602.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(Without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_20490_20590

Perfect score: 101
Sequence: 1 ctgcagctgagcagagctt.....ctggatgctgcgtcgcgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCtUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	71989	US-09-443-501A-2	Sequence 2, Appli
2	99.4	98.4	68750	US-09-335-409-1	Sequence 1, Appli
3	99.4	98.4	68750	US-09-568-102-1	Sequence 1, Appli
4	99.4	98.4	68750	US-09-567-969-1	Sequence 1, Appli
5	99.4	98.4	68750	US-09-568-480-1	Sequence 1, Appli
6	99.4	98.4	68750	US-09-568-486-1	Sequence 1, Appli
7	99.4	98.4	68750	US-09-568-472-1	Sequence 1, Appli
8	99.4	98.4	68750	US-09-567-899-1	Sequence 1, Appli
9	42.4	42.0	13842	US-09-105-537-30	Sequence 30, Appli
10	42.4	42.0	13842	US-08-804-227C-13	Sequence 13, Appli
11	42.4	42.0	36778	US-09-105-537-5	Sequence 5, Appli
12	42.4	42.0	38506	US-09-320-878-19	Sequence 19, Appli
13	42.4	42.0	43280	US-08-804-227C-1	Sequence 1, Appli
14	42.4	42.0	44377	US-08-804-227C-7	Sequence 7, Appli
15	42.4	42.0	44377	US-08-804-198-1	Sequence 1, Appli
16	40.8	40.4	11220	US-09-105-537-32	Sequence 32, Appli
17	39.8	39.4	4689	US-09-105-537-34	Sequence 34, Appli
18	39.2	38.8	1681	US-09-434-288-7	Sequence 7, Appli
19	39.2	38.8	20235	US-07-642-734C-3	Sequence 3, Appli
20	39.2	38.8	20235	US-08-439-009A-3	Sequence 3, Appli
21	38.2	37.8	8460	US-08-469-005A-9	Sequence 9, Appli
22	38.2	37.8	8519	US-09-261-907-1	Sequence 1, Appli
23	37.6	37.2	15872	US-09-105-537-1	Sequence 1, Appli
24	37.6	37.2	80161	US-09-036-987A-1	Sequence 1, Appli
25	37.6	37.2	80161	US-09-370-700-1	Sequence 1, Appli
26	37.4	37.0	33529	US-09-144-085-3	Sequence 3, Appli
27	36.4	36.0	11219	US-07-642-734C-1	Sequence 1, Appli

28	36.4	36.0	11219	US-08-439-009A-1	Sequence 1, Appli
c 29	34.2	33.9	748	US-09-154-083-1	Sequence 1, Appli
c 30	34.2	33.9	753	US-09-154-083-28	Sequence 28, Appli
c 31	33.4	33.1	50937	US-09-428-517-1	Sequence 1, Appli
c 32	33	32.7	4403765	US-09-103-840A-2	Sequence 2, Appli
c 33	32.8	32.5	444	US-09-060-756-658	Sequence 658, App
34	32.2	31.9	28958	US-08-256-261B-6	Sequence 6, Appli
35	32.2	31.9	28958	US-08-456-837-6	Sequence 6, Appli
36	32.2	31.9	28958	US-08-457-342-6	Sequence 6, Appli
37	32.2	31.9	28958	US-08-457-646A-6	Sequence 6, Appli
38	32.2	31.9	28958	US-08-458-076A-6	Sequence 6, Appli
39	32.2	31.9	28958	US-08-764-233A-4	Sequence 4, Appli
40	32.2	31.9	28958	US-08-457-335A-6	Sequence 6, Appli
41	32.2	31.9	28958	US-08-729-214-6	Sequence 6, Appli
42	32.2	31.9	28958	US-09-028-934-6	Sequence 6, Appli
43	32.2	31.9	49377	US-08-764-233A-1	Sequence 1, Appli
c 44	31.2	30.9	1434	US-09-434-288-3	Sequence 3, Appli
45	30.6	30.3	4041	US-09-105-537-36	Sequence 36, Appli

ALIGNMENTS

```
RESULT 1
US-09-443-501A-2
: Sequence 2, Application US/09443501A
: Patent No. 6303342
: GENERAL INFORMATION:
: APPLICANT: Kusan Biosciences, Inc.
: APPLICANT: Julien, Bryan
: APPLICANT: Katz, Leonard
: APPLICANT: Khosla, Chaitan
: APPLICANT: Tang, Li
: APPLICANT: Ziemann, Rainer
: TITLE OF INVENTION: Recombinant Methods and Materials for Producing
: TITLE OF INVENTION: Epothilone and Epothilone Derivatives
: FILE REFERENCE: 30062-20031.00
: CURRENT APPLICATION NUMBER: US/09/443,501A
: PRIOR FILING DATE: 1999-11-19
: PRIOR APPLICATION NUMBER: US 60/130,560
: PRIOR FILING DATE: 1999-04-22
: PRIOR APPLICATION NUMBER: US 60/122,620
: PRIOR FILING DATE: 1999-03-03
: PRIOR APPLICATION NUMBER: US 60/119,386
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: US 60/109,401
: PRIOR FILING DATE: 1998-11-20
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO. 2
: LENGTH: 71989
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match      100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 3.8e-22;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CCGCAGCGCGCAGCGAGCGCTGCGCAGCAGGCGCCGCTCGATGCGTGGAGATC 60
DB 20490  CTCGACGTCGCGCAGCGAGCGCTTCCCGCAGCAGGCGCTCGACTCTGATGCTGTGGAGATC 20549
QY      61  CGCAACGCGCTTCAGCGTGTAGCTGGGTATGCCGCTGTGCGGC 101
DB 20550  CGCAACGCGCTTCAGCGTGTAGCTGGGTATGCCGCTGTGCGGC 20590

RESULT 2
US-09-335-409-1
: Sequence 1, Application US/09335409
```


; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.2e-21;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 60
|||
Db 26102 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 26161
OY 61 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 101
|||
Db 26162 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 26202

RESULT 3
US-09-568-102-1

; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.2e-21;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 60
|||
Db 26102 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 26161
OY 61 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 101
|||
Db 26162 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 26202

RESULT 4
US-09-567-969-1

; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.2e-21;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 60
|||
Db 26102 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 26161
OY 61 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 101
|||
Db 26162 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 26202

RESULT 5
US-09-568-480-1

; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.2e-21;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 60
|||
Db 26102 CTCGACGTGGACGAGGCTTCGCCGAGCAAGGCTCGACTCCCTGATGGCTGTGAGATC 26161
OY 61 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 101
|||
Db 26162 CGCAACGGCTTCAGGCTGAGCTGGGTATCCGCTGTGCGC 26202


```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6353459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.2e-21;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 60
DB 26102 CTCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 26161

OY 61 CGCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 101
DB 26162 CGCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 26202

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.2e-21;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 60
DB 26102 CTCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 26161

OY 61 CGCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 101
DB 26162 CGCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 101
```

```
DB 26162 CGCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 26202

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 1.2e-21;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 60
DB 26102 CTCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 26161

OY 61 CGCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 101
DB 26162 CGCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 26202

RESULT 9
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match          42.0%; Score 42.4; DB 4; Length 13842;
Best Local Similarity 64.0%; Pred. No. 0.00027;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 2 TCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 61
DB 13286 TCGACGTGCAGCAGAGGCTTCGCCGAGCAGGAGCCCTCGACTCCCTGATGAGCTGTGGAGATC 13345

OY 62 GCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 101
DB 26162 GCAACGGCTTCAGAGGTGAGCTGGGTATGCCCGCTGTGGC 101
```

Db 13346 GCAACGGCTCAACTCCGCGGTGCGCTCGCCTCCCGGC 13385

RESULT 10

US-08-804-227C-13
; Sequence 13, Application US/08804227;
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13967 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
US-08-804-227C-13

Query Match

42.0%; Score 42.4; DB 2; Length 13987;
Best Local Similarity 64.0%; Pred. No. 0.00027;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 TCGACGTGCGAGCGCTTCGCCGAGAGTCCCTCGACTCCCGTGGATGCTGGAATCC 61

Db 13551 TCGAGCCCGCAACGACCCCTTCGGGAGATCGCTCGACTCCCGCGCGGTGAGCTGC 13610

QY 62 GCAACGGCTTCAGGCTGAGCTGGTATGCGCTGTCCGC 101

Db 13611 GCAACGGCTCAGCGCTGTGTCGGCTGCGTGGTGGCGAC 13650

RESULT 11

US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zuo, L.
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin
; FILE REFERENCE: 600,438051
; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 5

; LENGTH: 36778

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-105-537-5

Query Match

42.0%; Score 42.4; DB 4; Length 36778;

Best Local Similarity 64.0%; Pred. No. 0.00032;

Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 TCGACGTGCGAGCGCTTCGCCGAGAGTCCCTCGACTCCCGTGGATGCTGGAATCC 61

Db 15027 TCGACGCCGCAACGCGGCTTCCTCGACTCGGCTTCGACTCCCGAGCGCGCTGAACTCC 15086

QY 62 GCAACGGCTTCAGGCTGAGCTGGTATGCGCGCTGTCCGC 101

Db 15087 GCAACGGCTCAACTCCGCGGTGCGCTCGCCTCCCGGC 15126

RESULT 12

US-09-320-878-19

; Sequence 19, Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A

; CURRENT FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908

; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139

; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880

; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 19

; LENGTH: 38506

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-320-878-19

Query Match

42.0%; Score 42.4; DB 3; Length 38506;

Best Local Similarity 64.0%; Pred. No. 0.00032;

Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 TCGACGTGCGAGCGCTTCGCCGAGAGTCCCTCGACTCCCGTGGATGCTGGAATCC 61

Db 13169 TCGAGCCCGCAACGCGGCTTCCTCGACTCGGCTTCGACTCCCGAGCGCGCTGAACTCC 13228

QY 62 GCAACGGCTTCAGGCTGAGCTGGTATGCGCGCTGTCCGC 101

Db 13229 GCAACGGCTCAACTCCGCGGTGCGCTCGCCTCCCGGC 13268

RESULT 13

US-08-804-227C-1

; Sequence 1, Application US/08804227C

; Patent No. 5876991

```

GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuntzoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas' G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match          42.0%; Score 42.4; DB 2; Length 43280;
Best Local Similarity 64.0%; Pred. No. 0.00033;
Matches   64; Conservative    0; Mismatches   36; Indels      0; Gaps      0;

DY     2 TCGAGCTGCGACGAGCTTCCGCCGCAGGGGCTCGACTCCCTGTATGGCTGTGGAGATCC 61
||||| ||||| | | ||||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19512 TCGAGCTCGCGCGAGCTTCACAAGGAAGCGGGCTTGACTCCCTCAACCGCGCTGAACTGC 19571
||||| ||||| | | ||||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||

DY     62 GCACCAGGCTTCAGGGTGAAGCTGGGTTATGCCGCTGCGGC 101
||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cb 19572 GCACCAGGCTGCGCGCGCACCGGGGCTGAAGCTGCCCCGC 19611
||||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-08-804-227C-7
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuntzoss, Stuart A.
```

```

1 APPLICANT: Rosteck, Paul R., Jr.
2 APPLICANT: Sutton, Kimberly L.
3 TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
4 NUMBER OF SEQUENCES: 15
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: THOMAS G. PLANT 1501
7 STREET: LILLY CORPORATE CENTER
8 CITY: INDIANAPOLIS
9 STATE: IN
10 COUNTRY: USA
11 ZIP: 46285
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM Compatible
16 OPERATING SYSTEM: MS-DOS
17 SOFTWARE: ASCII(DOS) Text only
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/804.227C
21 FILING DATE: February 21, 1997
22
23 CLASSIFICATION: 435
24
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Plant, Thomas, G.
27 REGISTRATION NUMBER: 35,784
28 REFERENCE/DOCKET NUMBER: X-8231
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 317-276-2459
31
32 INFORMATION FOR SEQ ID NO: 7:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 44377 base pairs
35 TYPE: nucleic acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38
39 MOLECULE TYPE: DNA (genomic)
40
41 FEATURE:
42 NAME/KEY: CDS
43 LOCATION: 350..14002
44 FEATURE:
45 NAME/KEY: CDS
46 LOCATION: 14046..20036
47 FEATURE:
48 NAME/KEY: CDS
49 LOCATION: 20110..31284
50 FEATURE:
51 NAME/KEY: CDS
52 LOCATION: 31329..36071
53 FEATURE:
54 NAME/KEY: CDS
55 LOCATION: 36155..41830
56
57 US-08-804-227C-7
58
59 Query Match 42.0%; Score 42.4; DB 2: Length 44377;
60 Best Local Similarity 64.0%; Pred. No. 0.00033;
61 Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
62
63 QY 2 TCGAGCTGCGAGAGCTTCGCCGAGCAGAGGCGCTTGACTCCCTGATGCGTGTGAGATCC 61
64 DB 13566 TCGAGCCCGGAGACGACCTTCGCCGAGATCGGGGTCGACTCCCTGGCGGGTGGAGGATCG 13625
65
66 QY 62 GCAACGCGCTTCAGGGTGAAGCTGGGTATGCCGCTGTCCGC 101
67 DB 13626 GCAACCGGCGCTCAGCCGCTGTGTCGGCTGCGGCTTCCGAC 13665
68
69 RESULT 15
70 US-08-804-198-1
71 ; Sequence 1, Application US/08804198
72 ; Patent No. 5945320
73 ; GENERAL INFORMATION:
74 ; APPLICANT: Burgett, Stanley G.
75 ; APPLICANT: Kuhstoss, Stuart A.
76 ; APPLICANT: Rao, Nagaraaja R.
77 ; APPLICANT: Richardson, Mark A.
78 ; APPLICANT: Rosteck, Paul R., Jr.
79 ;

```

```

; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL, 1118
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
;
US-08-804-198-1

```

```

Query Match          42.0%; Score 42.4; DB 2; Length 44377;
Best Local Similarity 64.0%; Pred. No. 0.00033;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY      2  TCGACGTGCGACGACGCTTGCCGCGACAGCGCTGCACTCCGTGATGCGCTGTGGAGATCC 61
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13566 TCGAGCCCGAAGCACCTTCGGGAGATCGCGTGCACCTCCCTGGCGCGGTGGAGCTGC 13625

QY      62  GCAAGCGGCTTCAAGCGCTGAGCTGGCTATGCGCTGTGCGC 101
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13626 GCAACCGGCTCAGCGCTGTGCGGCTGCGGTGCGCGAC 13665

```

Search completed: November 5, 2002, 14:07:34
Job time : 171.737 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_20490_20590

Perfect score: 101

Sequence: 1 ctgcagctgacgagcgtt.....ctggatgacgctgctgcgc 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCR_NEM_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.4	42.0	13842	10	US-09-861-289-30
2	42.4	42.0	36778	10	US-09-861-289-5
3	40.8	40.4	11220	10	US-09-861-289-32
4	39.8	39.4	4689	10	US-09-861-289-34
5	38.2	37.8	3209	10	US-09-925-301-474
6	37.6	37.2	15872	10	US-09-861-289-1
7	30.6	30.3	4041	10	US-09-861-289-36
8	29.4	29.1	1083	10	US-09-815-242-7578
9	29.4	29.1	3189	12	US-10-044-090-333
10	29.4	29.1	3299	10	US-09-800-729-68
11	29.4	29.1	6065	10	US-09-800-729-35
12	28.8	28.5	284	10	US-09-294-093B-4947
13	28.4	28.1	1830	10	US-09-731-872-241
14	28.4	28.1	1831	10	US-09-731-872-240
15	28.4	28.1	2492	10	US-09-925-301-593
16	27	26.7	884	10	US-09-833-381-1048
17	27	26.7	3132	10	US-09-815-242-1139
18	27	26.7	167343	10	US-09-962-436-281
19	27	26.7	167343	10	US-09-964-824A-273

C 20	26.8	26.5	1161	10	US-09-808-387-5	Sequence 5, Appl
C 21	26.8	26.5	1674	10	US-09-808-387-3	Sequence 3, Appl
C 22	26.8	26.5	1725	10	US-09-808-387-1	Sequence 1, Appl
C 23	26.4	26.1	2298	10	US-09-866-582-21	Sequence 21, Appl
C 24	26	25.7	272	10	US-09-878-574-5740	Sequence 5740, Ap
C 25	26	25.7	577	10	US-09-864-761-9695	Sequence 9695, Ap
C 26	26	25.7	1023	10	US-09-814-777A-16	Sequence 16, Appl
C 27	26	25.7	1155	10	US-09-814-777A-21	Sequence 21, Appl
C 28	26	25.7	1421	10	US-09-814-777A-14	Sequence 14, Appl
C 29	26	25.7	1730	10	US-09-814-777A-19	Sequence 19, Appl
C 30	26	25.7	1919	10	US-09-814-777A-17	Sequence 17, Appl
C 31	25.8	25.5	257	10	US-09-923-876-3033	Sequence 3033, Ap
C 32	25.8	25.5	3186	10	US-09-880-107-1659	Sequence 1659, Ap
C 33	25.8	25.5	4358	10	US-09-917-800A-1343	Sequence 1343, Ap
C 34	25.8	25.5	13029	10	US-09-815-242-4052	Sequence 4052, Ap
C 35	25.6	25.3	1888	10	US-09-925-300-722	Sequence 722, App
C 36	25.6	25.3	2001	10	US-09-768-877-7	Sequence 7, Appl
C 37	25.6	25.3	2516	10	US-09-768-877-11	Sequence 11, Appl
C 38	25.6	25.3	3177	10	US-09-898-570-15	Sequence 15, Appl
C 39	25.6	25.3	35871	10	US-09-956-335-2	Sequence 2, Appl
C 40	25.6	25.3	35935	10	US-09-725-720-43	Sequence 43, Appl
C 41	25.6	25.3	35935	10	US-09-782-378A-4	Sequence 4, Appl
C 42	25.6	25.3	35935	10	US-09-782-378A-5	Sequence 5, Appl
C 43	25.6	25.3	35978	10	US-09-956-335-1	Sequence 1, Appl
C 44	25.4	25.1	147	10	US-09-864-761-24022	Sequence 24022, A
C 45	25.4	25.1	487	10	US-09-833-381-1125	Sequence 1125, Ap

ALIGNMENTS

```
RESULT 1
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patient No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match      42.0%  Score 42.4:  DB 10:  Length 13842:
Best Local Similarity 64.0%:  Pred. No. 0.00016:
Matches 64:  Conservative 0:  Mismatches 36:  Indels 0:  Gaps 0:

QY      2 TCAGCTGTCGACGAGCTTCGCCGACGACGAGCTGCCTGATGCTGTGAGATCC 61
        |||||  ||| ||||| ||| ||| ||||| ||||| ||| ||| |||
DB 13286 TCAGCGCCGACGAGGGGCTTCCTCGACCTCGGCTGTGACCTCGACCGCGCGAATCC 13345

QY      62 GCAACGCTTCAGGCTGAGCTGAGCTGATCCCTGTCGCGC 101
        ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13346 GCAACGCTTCAGGCTGAGCTGAGCTGAGCTGATCCCTGTCGCGC 13385

RESULT 2
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patient No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
```

```

; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/851,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-5
```

```
Query Match          42.0%; Score 42.4; DB 10; Length 36778;
Best Local Similarity 64.0%; Pred No. 0.00019;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
QY 2 TCACAGCTGCGACGAGCTTCGCCGAGCAGCGCTCGACTCCCTGATGCTGTGGAGATCC 61
    ||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 15027 TCACAGCGCGAAGCGGGCTTCCTCGACTCGCTCGACTCGCTGACCGCGCTGCACTCC 15086
QY 62 GCAACGCGCTTCAGAGGTGAGCTGGGTATCCGCTGTGCGC 101
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15087 GCAACGCGCTCAACTCCGCGGTGGCTTCGCCCTCCGCGC 15126
```

```
RESULT 3
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-32
```

```
Query Match          40.4%; Score 40.8; DB 10; Length 11220;
Best Local Similarity 63.0%; Pred. No. 0.00049;
Matches 63; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 2 TCACAGTGCACGAGGCTTCGCCGAGCAGCGCTCGACTCCCTGATGCTGTGGAGATCC 61
    ||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 4400 TCACACGCGGGCGGGCTTCGCGACTCGATTCGACTCGCTGCGAGCGGCTGAGAGCTCC 4459
QY 62 GCAACGCGCTTCAGAGGTGAGCTGGGTATCCGCTGTGCGC 101
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4460 GCAACGCGCTCAAGAACGCCACCGGCTGTCCTCCCGGC 4499
```

```
RESULT 4
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
```

```

; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-34
```

```
Query Match          39.4%; Score 39.8; DB 10; Length 4689;
Best Local Similarity 64.8%; Pred. No. 0.00086;
Matches 59; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
QY 11 GACGAGCTTCGCCGAGCAGCGCTTCGACTCCCTGATGCTGTGGAGATCCGCAACGCC 70
    ||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 4283 GCCGTGCTTACCGAGCTTCGCTGCTGAGCTGAGCGCGCTGCGCAACGAGC 4342
QY 71 TTCAGGCTGAGCTGGGTATCCGCTGTGCGC 101
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4343 TCACAGCGTGTGCGGCAACGAGCTCCCGC 4373
```

```
RESULT 5
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 474
; LENGTH: 3209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (427)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-301-474
```

```
Query Match          37.8%; Score 38.2; DB 10; Length 3209;
Best Local Similarity 61.6%; Pred. No. 0.0025;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 2 TCACGTGGAGCAGGCTTCGCCGAGCAGCGCTCGACTCCCTGATGCTGTGGAGATCC 61
    ||||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| |||||
Db 1263 TCACCTGACACACTGACTGCGGAGCTTGAGCTTCGACTGCTCATGAGCTGGAGGTCC 1322
QY 62 GCAACGCGCTTCAGAGGTGAGCTGGGTATCCGCTGTGCGC 100
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1323 GCCAGACGCTGAGCGTGAAGCTCAACTGTGTGCTGCC 1361
```

```
RESULT 6
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
```

```

1  APPLICANT: Sherman, D.H.
2  APPLICANT: Liu, H.
3  APPLICANT: Xue, Y.
4  APPLICANT: Zhao, L.
5  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
6  FILE REFERENCE: 600,438US1
7  CURRENT APPLICATION NUMBER: US/09/861,289
8  CURRENT FILING DATE: 2001-05-18
9  PRIOR APPLICATION NUMBER: 09/105,537
10 PRIOR FILING DATE: 1998-06-26
11 NUMBER OF SEQ ID NOS: 43
12 SOFTWARE: FASTSEQ for Windows Version 3.0
13 SEQ ID NO 1
14     LENGTH: 15872
15     TYPE: DNA
16 ORGANISM: Streptomyces venezuelae
17 US-09-861-289-1

```

Query Match	37.2%;	Score 37.6;	DB 10;	Length 15872;
Best Local Similarity	61.0%;	Pred. No. 0.0051;		
Matches 61; Conservative	0;	Mismatches 39;	Indels 0;	Gaps 0;

DY 2 TCGACGTGCGAGCAGGGCTTGC GCCAGCACGGCC TCACATCCCTGATGSGCTGTGGAGATCC 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8253 TCGGGGCGGAGCGGTCCTTCAGGACTCGGCTTGCACTCCCTGGCCGGGGTTGGAGATCC 8312

OY 62 GCAACGGCTTCAGGTTGACTGTGGTATGCCCGTGTCGGC 101
 ||| | | | | | | | | | | | | |
Dd 8313 GCAACGGGCTGAACGGCCGCACACGGGCTGGGGCTCCCCGC 8352

```

US-09-861-289-36
US-09 7
: Sequence 36, Application US/09861289
: Patent No. US20020110897A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D. H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and plikromycin
: FILE REFERENCE: 600.4380S1
: CURRENT APPLICATION NUMBER: US/09/861,289
: CURRENT FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 36
: LENGTH: 4041
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-861-289-36

```

	Query Match	Score 30.3%	DB 10	Length 4041
	Best Local Similarity	62.3%	Pred. No. 0.62	
Matches	48	Conservative	0	Mismatches 29; Indels 0; Gaps 0;
OY	25	GAGCAGGGCCGATGATCTCCTGATGCGCTGTGGAGATGCCCAACGCGCTTCAGGGTAAGCTG	84	
DB	2923	GAGATCGGCTTGATGATCTGCGTCGACCGCGCTGCACTTCGCGAACCGGCTCAACCGG.TGACC	2982	
OY	85	GGATGCGCGCTGTGCGC	101	
DB	2983	GCTCTCCAGCTGCGCC	2999	

RESULT 8
US-09-815-242-7578/c
; Sequence 7578, Application US/09815242;
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

```

1  APPLICANT: Ohlsen, Karl L.
2  APPLICANT: zyskind, Judith W.
3  APPLICANT: Wall, Daniel
4  APPLICANT: Trawick, John D.
5  APPLICANT: Carr, Grant J.
6  APPLICANT: Yamamoto, Robert T.
7  APPLICANT: Xu, H. Howard
8  TITLE OF INVENTION: Identification of Essential Genes in
9  TITLE OF INVENTION: Prokaryotes
10 FILE REFERENCE: ELITRA.011A
11 CURRENT APPLICATION NUMBER: US/09/815,242
12 CURRENT FILING DATE: 2001-03-21
13 PRIOR APPLICATION NUMBER: 60/191,078
14 PRIOR FILING DATE: 2000-03-21
15 PRIOR APPLICATION NUMBER: 60/206,848
16 PRIOR FILING DATE: 2000-05-23
17 PRIOR APPLICATION NUMBER: 60/207,727
18 PRIOR FILING DATE: 2000-05-26
19 PRIOR APPLICATION NUMBER: 60/242,578
20 PRIOR FILING DATE: 2000-10-23
21 PRIOR APPLICATION NUMBER: 60/253,625
22 PRIOR FILING DATE: 2000-11-27
23 PRIOR APPLICATION NUMBER: 60/257,931
24 PRIOR FILING DATE: 2000-12-22
25 PRIOR APPLICATION NUMBER: 60/269,308
26 PRIOR FILING DATE: 2001-02-16
27 NUMBER OF SEQ ID NOS: 14110
28 SOFTWARE: FastSeq for Windows Version 4.0
29 SEQ ID NO 7578
30 LENGTH: 1083
31 TYPE: DNA
32 ORGANISM: Klebsiella pneumoniae
33 FEATURE:
34 NAME/KEY: CDS
35 LOCATION: (1)...(1083)
36 OS-09-815-242-7578

```

Query Match	29.1%	Score 29.4	DB 10	Length 1083
Best Local Similarity	63.4%	Pred. No. 1.2		
Matches	45	Conservative	0	Mismatches 26; Indels 0; Gaps 0;
QY	2	TCGAGCTGCAGAGAGCTTCGGCCGAGCAAGGGCGCTGCATCCCTGATGGCTGTGAGATCC	61	
Db	92	TCGGCAATCGTCGGCGCATCGCCGAGCAGCGCTGCATCTTCATGGCGCTGTACAAC	33	
QY	62	GCAACGGCTT	72	
Db	32	GCTTCAGATT	22	

```

RESULT 9
US-10-044-090-333
: Sequence 333, Application US/10044090
: Patent No. US20020137081A1
: GENERAL INFORMATION:
: APPLICANT: Olga Bandman
: TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
: FILE REFERENCE: PA-0028 US
: CURRENT APPLICATION NUMBER: US/10/044,090
: CURRENT FILING DATE: 2002-01-09
: NUMBER OF SEQ ID NOS: 850
: SOFTWARE: PERL Program
: SEQ ID NO 333
: LENGTH: 3189
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No. US20020137081A1 2960079CB1
US-10-044-090-333

```

Query Match	29.1%;	Score 29.4;	DB 12;	Length 3189,
Best Local Similarity	58.68;	Pred. No. 1.4;		

Matches	51: Conservative	0: Mismatches	36: Indels	0: Gaps
0y	10 CGACGAGCCTTGGCCGAGAGGCGCTTCGACTCCCTGATGCTGTGGAGATCCGCACAGG	69		
Db	179 CCAACGGGGGGGAGAGAGAGAGGCCACGATCTCTCCGGCGCGGAGAGAGCTACGACACGCTG	238		
0y	70 CTTCAGGAGACTGGGATATGCCGCTG	96		
Db	239 GCGCGCGCCGACAGTCGAGTCGTCCGCTG	265		

```

US-09-800-729-68
RESULT 10
US-09-800-729-68
Sequence 68, Application US/05800729
Patent No. US20020068319A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 32 Human secreted proteins
FILE REFERENCE: P044P1
CURRENT APPLICATION NUMBER: US/09/800,729
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: PCT/US00,426013
PRIOR FILING DATE: 2000-09-22
PRIOR FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 217
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 68
LENGTH: 3299
TYPE: DNA
ORGANISM: Homo sapiens
US-09-800-729-68

```

Query Match	29.1%	Score: 29.4	DB 10	Length 3299
Best Local Similarity	58.6%	Pred. No. 1.4		
Matches 51	Conservative 0	Mismatches 36	Indels 0	Gaps 0

Qy	Dy	Dx	Qx
10	CGACGAGGCTTCGGCCGAGACAGGGCTCGPCTCCCTGATGGCTGGAGATCCGAAACGG	69	
249	CCAAAGCGCGGGGAGAAAGACAGAGCCCAATCTGCCCCCGCGCGGAGAGCTACGAACTG	308	
70	CTTCAGGCTGAGCTGGGTATCCGCTG	96	
309	CGCCGCGCCGACGTCGGCTGTCGCGCTG	335	

```

RESULT 11
US-09-800-729-35
Sequence 35, US/09800729
Patent No. US20020068319A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 32 Human secreted pr
FILE REFERENCE: P/044p1
CURRENT APPLICATION NUMBER: US/09/803,729
CURRENT FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: PCT/US00/036013
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/155,703
PRIOR FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 217
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 35
LENGTH: 6065
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (6035)
OTHER INFORMATION: n equals a,t,g, c
NAME/KEY: SITE
LOCATION: (6037)
OTHER INFORMATION: n equals a,t,g, c

```

```

; NAME/KEY: SITE
; LOCATION: (6038)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-800-729-35

```

Query Match	29.18;	Score 29.4;	DB 10;	Length 6065;
Best Local Similarity	58.68;	Pred. No. 1.6;		
Matches 51;	Conservative	0;	Mismatches 36;	Indels 0;
				Gaps 0

Oy	10	CGACGAGGCTTTCGGCGAGACGGGCTCGACATCCCTGATGGCTGTGGAGATCCGCAACGG	69
Dd	249	CCAAACGGCGGGGAAGACGACGACGCCCACTATGCCGGGGGGAGAGAGCTACGCAACTG	308
Oy	70	CTTCAGGGGTGAGCTGGGTATGCCCGTG	96
Dd	309	CGCGCGCGCCGACGTCGATCGTCCCGTG	335

RESULT 12
US-09-294-093B-4947

; Sequence 4947, Application US/09294093EE
; Patent No. US20010051335A1
; GENERAL INFORMATION:

;; GENERAL INFORMATION:
;; APPLICANT: Lalgudi, Raghunath, V.
;; APPLICANT: Ito, Laura, Y.

```

; APPLICANT:  Sherman, Bradley, K.
; TITLE OF INVENTION:  POLYNUCLEOTIDES A

```

```

; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093E
CURRENT FILING DATE: 1000-04-15

```

;; CURRENT FILING DATE: 1999-04-18
;; PRIOR APPLICATION NUMBER: 60/082,567
;; PRIOR FILING DATE: April 21, 1998

```

; NUMBER OF SEQ ID NOS:
; SOFTWARE: PERL Program
;

```

```

; SEQ ID NO 4947
; LENGTH: 284

```

```

: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:

```

```

/
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700355442H1
;

```

Query Match	28.5%	Score 28.8	DB 10	Length 284
Best Local Similarity	60.0%	Pred No. 1.4		
Matches 48	Conservative 0	Mismatches 32	Indels 0	Gaps 0

Oy 2 TCGAGCTGGCAGCAGGCGTTCCGCCGACAGGGCCCTGCACCTCCCCTGATGGCTGTGGAGATCC 61
 ||||| - ||| ||| ||||| ||| | |||||||||
Db 34 TCGACGACTGGAGAGAGAGCACACGAGAAGAGCGTCGCTTTTGGCCGACGTGTGGAGATCA 93

QY	62	GCAACGGCTTCAGGGTGA	81
Db	94	TCAACGGCGACGGCGGAG	113

RESULT 13
US-09-731-872-241

```

; Sequence 241, Application US/09/31872
; Patent No. US20020102604A1
: GENERAL INFORMATION:

```

```

/ GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bouqueleret, Lydie

```

APPLICANT: Robert, Severin
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS,

```

; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07

```

;; CURRENT FILING DATE: 2000-12-07
;; PRIOR APPLICATION NUMBER: US 60/169,629
;; PRIOR FILING DATE: 1999-12-08

PRIOR APPLICATION NUMBER: US 60/187,470
PRIOR FILING DATE: 2000-03-06

; NUMBER OF SEQ ID NOS: 482

SOFTWARE: Patent.pm
SEQ ID NO 241
LENGTH: 1830
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 78..608
US-09-731-872-241

Query Match 28.1%; Score 28.4; DB 10; Length 1830;
Best Local Similarity 58.1%; Pred. No. 2.6;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 16 GGCTTCGCCGAGCAGGCGCTCGACTCCCGATGCTGTGAGATCCGCAAGCGCTTCAG 75
DB 683 GTCTTCGGGGAGCTGTGGGGCGAAGCTGCCGCTGTGGGGCCCTGATGAGCGCTTC 742

QY 76 GGTGAGCTGGGTATGCGCGCTGTCGCGC 101
DB 743 GGTGTGCTGTGAGCGCTGTGTGTTC 768

RESULT 14
US-09-731-872-240

Sequence 240; Application US/09731872
Patent No. US20020102604A1

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, Jean Baptiste

APPLICANT: Bouquelere, Lydie

APPLICANT: Jobert, Severin

TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS

FILE REFERENCE: 78. US3. REG

CURRENT APPLICATION NUMBER: US/09/731.872

CURRENT FILING DATE: 2000-12-07

PRIOR APPLICATION NUMBER: US 60/169,629

PRIOR FILING DATE: 1999-12-08

PRIOR APPLICATION NUMBER: US 60/187,470

PRIOR FILING DATE: 2000-03-06

NUMBER OF SEQ ID NOS: 482

SOFTWARE: Patent.pm

SEQ ID NO 240

LENGTH: 1831

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: 78..590

US-09-731-872-240

Query Match 28.1%; Score 28.4; DB 10; Length 1831;
Best Local Similarity 58.1%; Pred. No. 2.6;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 16 GGCTTCGCCGAGCAGGCGCTCGACTCCCGATGCTGTGAGATCCGCAAGCGCTTCAG 75
DB 684 GTCTTCGGGGAGCTGTGGGGCGAAGCTGCCGCTGTGGGGCCCTGATGAGCGCTTC 743

QY 76 GGTGAGCTGGGTATGCGCGCTGTCGCGC 101
DB 744 GGTGTGCTGTGAGCGCTGTGTGTTC 769

RESULT 15
US-09-925-301-593

Sequence 593; Application US/09925301
Patent No. US20020052308A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA106

CURRENT APPLICATION NUMBER: US/09/925,301

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05882

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1694

SOFTWARE: Patent Ver. 2.0

SEQ ID NO 593

LENGTH: 2492

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

LOCATION: (2113)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc.feature

LOCATION: (2452)

OTHER INFORMATION: n equals a,t,g, or c

US-09-925-301-593

Query Match 28.1%; Score 28.4; DB 10; Length 2492;
Best Local Similarity 58.1%; Pred. No. 2.7;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 16 GGCTTCGCCGAGCAGGCGCTCGACTCCCGATGCTGTGAGATCCGCAAGCGCTTCAG 75
DB 671 GTCTTCGGGGAGCTGTGGGGCGAAGCTGCCGCTGTGGGGCCCTGATGAGCGCTTC 730

QY 76 GGTGAGCTGGGTATGCGCGCTGTCGCGC 101
DB 731 GGTGTGCTGTGAGCGCTGTGTGTTC 756

Search completed: November 5, 2002, 22:59:50
Job time : 97.2537 secs


```
FT      /*tag- d
FT      /note- "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /*tag- e
FT      /note- "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /*tag- f
FT      /note- "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /*tag- g
FT      /note- "encodes a xyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /*tag- h
FT      /label- epoc_gene
FT      /note- "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /*tag- i
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /*tag- j
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /*tag- k
FT      /note- "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /*tag- l
FT      /note- "encodes c:condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /*tag- m
FT      /note- "encodes c:condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag- n
FT      /note- "encodes acenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag- o
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /*tag- p
FT      /note- "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /*tag- q
FT      /note- "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /*tag- r
FT      /note- "encodes acenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /*tag- s
FT      /note- "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /*tag- t
FT      /note- "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /*tag- u
FT      /note- "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /*tag- v
FT      /note- "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /*tag- w
FT      /note- "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /*tag- x
```

```
FT      /label- epoc_gene
FT      /note- "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /*tag- y
FT      /note- "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /*tag- z
FT      /note- "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /*tag- aa
FT      /note- "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /*tag- ab
FT      /note- "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /*tag- ac
FT      /note- "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /*tag- ad
FT      /label- epod_gene
FT      /note- "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /*tag- ae
FT      /note- "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /*tag- af
FT      /note- "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /*tag- ag
FT      /note- "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /*tag- ah
FT      /note- "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /*tag- ai
FT      /note- "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /*tag- aj
FT      /note- "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /*tag- ak
FT      /note- "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /*tag- al
FT      /note- "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /*tag- am
FT      /note- "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /*tag- an
FT      /note- "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /*tag- ao
FT      /note- "encodes DH5"
FT      misc_RNA
FT      29433..30287
FT      /*tag- ap
FT      /note- "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /*tag- aq
FT      /note- "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /*tag- ar
FT      /note- "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /*tag- as
FT      /note- "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /*tag- at
FT      /note- "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /*tag- au
```


XX NPI: 2000-097741/08.
DR AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR P-PSDB: AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR AAY58592, AAY58593, AAY58594.
XX
PT New isolated epothilone synthase genes, used for the recombinant
PT production of epothilone for use in cancer therapy -
XX
PS Claim 14: Page 87-104; 174pp; English.

CC This sequence represents a 68.75 kb contig from *Sorangium cellulosum*
CC comprising 22 open reading frames (ORFs) and includes genes encoding
CC proteins involved in the biosynthesis of epothilones. Epothilones A and
CC B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC the chlaetole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AAY58575-58578) are involved in polyketide backbone
CC formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epothilones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epothilones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epothilones as anticancer agents, they are problematical to produce on a
CC large scale. Epothilones are too complex for industrial scale chemical
CC synthesis, and *Sorangium cellulosum* is difficult to ferment, producing
CC poor yields of epothilones. The nucleic acids of the invention may be
CC used for the recombinant production of epothilones in a heterologous host
CC that is more amenable to fermentation.
XX

XX S0 Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other:

Query Match 98.4%; Score: 99.4; DB 21; Length 68750;
Best Local Similarity 99.0%; Pred. No. 7.4e-20;

Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGACGCTGCGACGAGCTTCGCGAGCAGGCTCTGACCTCCGATGCGTGTGAGATC 60
|||||

DB 26102 CTGACGCTGCGACGAGCTTCGCGAGCAGGCTCTGACCTCCGATGCGCGTGTGAGATC 26161
|||||

QY 61 CGCAACGGCTTCAGGCTGAGCTGAGTATGCGCTGTGCGC 101
|||||

DB 26162 CGCAACGGCTTCAGGCTGAGCTGAGTATGCGCTGTGCGC 26202
|||||

RESULT 3

AAF24892 ID AAF24892 standard; DNA; 20394 BP.

XX AAF24892;

XX 20-APR-2001 (first entry)

DE Pimaricin biosynthesis associated polyketide synthase gene.

XX Polyketide synthase; oxidative modification; metabolite; antidiabetic;
KW anticancer; pimaricin; ss.

XX Streptomyces natalensis.

XX Key Location/Qualifiers:
FH 1..20394
FT /tag- a
FT /product- "polyketide synthase"

XX

PN M0200077222-A1.

XX 21-DEC-2000.

XX 14-JUN-2000; 2000MO-EP06227.

XX 14-JUN-1999; 99EP-0201893.

XX (STAM) DSM NV.

PI Martin JF, Aparicio JF, Colina AJ.

DR WPI: 2001-080693/09.

DR P-PSDB: AAB31558.

PT New polyketide synthase genes encoding enzymes involved in the biosynthesis of
PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT the synthesis of new compounds -
XX

PS Disclosure; Page 53-80; 116pp; English.

CC The present sequence encodes a polyketide synthase which is associated
CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC is useful for the oxidative modification of a methyl group of a suitable
CC compound, e.g. a bioactive compound including a secondary metabolite,
CC antibiotics and anticancer agents. Recombinant cells comprising the
CC gene are useful for the production of pimaricin. The polyketide synthase
CC polynucleotide may be over expressed in *Streptomyces*, leading to an
CC increase in the biosynthesis of pimaricin, as a source of primers for
CC amplification reaction and as probes.
XX

XX S0 Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other:

Query Match 51.5%; Score 52; DB 22; Length 20394;
Best Local Similarity 70.0%; Pred. No. 6e-06;

Matches 70; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 2 TCGACGCTGCGACGAGCTTCGCGAGCAGGCTTCGCGATGCGTGTGAGATC 61
|||||

DB 4469 TCGATGCCCAACGAGCGCTTCGCGAGCAGGCTTCGCGATGCGTGTGAGATC 4528
|||||

QY 62 GCAACCGCTTCAGGCTGAGCTGAGTATGCGCTGTGCGC 101
|||||

DB 4529 GCAACCGCTTCGCGAGCAGGCTTCGCGAGCAGGCTGTGCGC 4568
|||||

RESULT 4

AAA92301 ID AAA92301 standard; DNA; 30690 BP.

XX AAA92301;

XX 10-JAN-2001 (first entry)

DE *S. avermitilis* avermectin aglycon synthase DNA avai seq ID NO:1.

XX Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.

XX Streptomyces avermitilis.

XX Key Location/Qualifiers

XX CDS 1..11919

XX /tag- a
FT /note- "avermectin aglycon synthase protein"

FT 11971..30690

FT /tag- b
FT /note- "avermectin aglycon synthase protein"

PN M0200050605-A1.

XX 31-AUG-2000.

XX 23-FEB-2000: 2000WO-JP01041.
PF
XX
PR 24-FEB-1999: 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI: 2000-565458/52.
DR P-PSDB: AAB23749, AAB23750.
XX
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 2: Page 66-134; 314pp; Japanese.
XX
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
XX
SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;
XX
Query Match 43.6%; Score 44; DB 21; Length 30690;
Best Local Similarity 65.0%; Pred. No. 0.0014;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
XX
QY 2 TCACGCTGCAGCAGGCTTCGCCGACGAGGCGCTCGACTCCCTGATGCGTGTGAGATCC 61
Db 30161 TCGACGCATCGCGGGGCTTCGCCGAGCTGGGCTTCGACTCGTCACGCGCGGTGAGACTGC 30220
QY 62 GCAACGGCTTCAGGGTGAAGCTGGTATGCCGCTGTCGGC 101
Db 30221 GCAACCGCTTCACGCGCGCGGCGCTGCGCTGCGGGC 30260
XX
RESULT 5
AAH79277
ID AAH79277 standard; DNA; 30690 BP.
XX
AC AAH79277;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermectilis coding sequences SEQ ID NO: 1.
XX
KM Avermectin aglycone synthase; AAS: avermectin derivative;
KM drug production; veterinary drug; pesticide; ds.
XX
OS Streptomyces avermectilis.
XX
FH Key Location/Qualifiers
FT CDS 1..11919
FT FT /tag= a
FT FT /product= "AAG65264"
FT FT /partial
FT FT 11971..30690
FT FT /tag= b
FT FT /product= "AAG65265"
XX
XX MO200162939-A1.
XX
PD 30-AUG-2001.
XX

PF 23-FEB-2001: 2001WO-JP01381.
XX
XX 24-FEB-2000: 2000JP-0047405.
XX
XX (KYOW) KYOMA HAKKO KOGYO KK.
XX
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI: 2001-582053/65.
DR P-PSDB: AAG65264, AAG65265.
XX
XX New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Example 2: Page 58-123; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermectilis genome.
XX
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
XX
Query Match 43.6%; Score 44; DB 22; Length 30690;
Best Local Similarity 65.0%; Pred. No. 0.0014;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
XX
QY 2 TCACGCTGCAGCAGGCTTCGCCGACGAGGCGCTCGACTCCCTGATGCGTGTGAGATCC 61
Db 30161 TCGACGCATCGCGGGGCTTCGCCGAGCTGGGCTTCGACTCGTCACGCGCGGTGAGACTGC 30220
QY 62 GCAACGGCTTCAGGGTGAAGCTGGTATGCCGCTGTCGGC 101
Db 30221 GCAACCGCTTCACGCGCGCGGCGCTGCGCTGCGGGC 30260
XX
RESULT 6
AAD17184
ID AAD17184 standard; DNA; 65140 BP.
XX
AC AAD17184;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nysl DNA of nystatin PKS gene cluster.
XX
KM Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KM antifungal; antibiotic; nysl; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (1..1035)
FT FT /tag= a
FT FT /product= "NysD2 partial protein"
FT FT /note= "CDS does not include stop codon"
FT FT complement (1056..2576)
FT FT /tag= b
FT FT /product= "NysD1 protein"
FT FT 2806..6906
FT FT /tag= c
FT FT /product= "NysA protein"
FT FT 6952..16530
FT FT /tag= d
FT FT /product= "NysB protein"
FT FT 16550..49840
CDS

FT	/tag- e
FT	/product= "NysC protein"
FT	50260..51015
FT	//tag- f
FT	/product= "NysE protein"
FT	51405..54305
FT	//tag- g
FT	/product= "NysR1 protein"
FT	54329..57190
FT	//tag- h
FT	/product= "NysR2 protein"
FT	/note= "CDS does not include start codon"
FT	57180..59963
FT	//tag- i
FT	/product= "NysR3 protein"
FT	60415..61047
FT	//tag- j
FT	/product= "NysR4 (short) protein"
FT	/note= "CDS does not include start codon"
FT	61736..62497
FT	//tag- k
FT	/product= "NysR5 protein"
FT	/note= "CDS does not include start codon"
FT	complement (62551..63615)
FT	//tag- l
FT	/product= "ORF2 protein"
FT	/note= "CDS does not include start codon"
FT	63765..64961
FT	//tag- m
FT	/product= "ORF1 protein"
XX	
PX	MO200159126-A2.
XX	
PD	16-AUG-2001.
PE	08-FEB-2001; 2001WC-GH00509.
XX	
PR	08-FEB-2000; 2000GB-0002840.
PR	10-APR-2000; 2000GB-0008786.
PR	14-APR-2000; 2000GB-0009387.
XX	
PA	(UNO-) UNIV NORGES TEKNISK NATUR/TITENSKAPELIGE.
PA	(SNPF) SINTEF STIPELSEN IND TEK FORSK.
PA	(ALPH-) ALPHARMA AS.
PA	(SINV-) SINVENT AS.
PA	(DATE/) DEIGLEMSKA H.
PA	(ZOTC/) ZOTCHEV S.B.
PA	(SEKU/) SEKUROVA O.N.
PA	(FUAE/) FJAEVRVIK E.
PA	(BRAU/) BRAUTASET T.
PA	(STRO/) STROM A.R.
XX	
PI	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI	Valla S, Ellingsen TE, Sletta H, Gulliksen O,
XX	
DR	WPI: 2001-557614/62.
DR	P-PSD; AAEE10123, AAEE10126, AAEE10127, AAEE10128, AAEE10129, AAEE10130,
XX	AAEE10131, AAEE10132, AAEE10133, AAEE10134, AAEE10135, AAEE10136, AAEE10137.
XX	
PT	New nystatin polyketide synthase polynucleotides and polypeptides,
PS	useful as antibiotics -
XX	
PS	Claim 2; Page 116-151; 266pp; English.
CC	The present invention relates to the cloning and sequencing of the gene
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC	involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC	The nystatin PKS is useful as antifungal antibiotics. The present
CC	sequence is a Streptomyces noursei nyl1 DNA of nystatin PKS gene cluster
XX	
SQ	Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

Query Match 43.6% Score 44 DB 22 Length 65140;

Best Local Similarity	65.0%	Pred. No. 0.0016;	Matches 65;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;
QY	2	TCGACGTCCGACGAGGCTTCGCCGACGACGAGCCCTGCATCCCTGATGCTGTGAGATCC	61				
Db	49425	TCGGTCCGGGGGAGACAGTTCGAGGATGCGGTTCGACATCGCTGACCCGGGTGAGATTCC	49484				
QY	62	GCAACGCGCTTCAGGCTGAGCTGGGTATGCGCGCTGCGGC	101				
Db	49485	GCAACGCACTCAACGCGCGCACCGGACTCGGCTGCGGC	49524				
RESULT 7							
XX	AD17186	standard; DNA; 125401 BP.					
XX	AD17186;						
XX	29-NOV-2001	(first entry)					
DE	Streptomyces noursei	PKS gene cluster DNA.					
KM	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;						
KW	antifungal; antibiotic; ds.						
OS	Streptomyces noursei.						
PH	Key	Location/Qualifiers					
FT	CDS	6337..34771					
FT		/*tag= a					
FT		/*product= "NysI complete protein"					
FT	CDS	34792..51099					
FT		/*tag= b					
FT		/*product= "NysJ protein"					
FT	CDS	51155..57355					
FT		/*tag= c					
FT		/*product= "NysK protein"					
FT	CDS	57503..58687					
FT		/*tag= d					
FT		/*product= "NysL protein"					
FT	CDS	complement (58786..58980)					
FT		/*tag= e					
FT		/*product= "NysM protein"					
FT		/note= "CDS does not include start codon"					
FT	CDS	complement (59045..60241)					
FT		/*tag= f					
FT		/*product= "NysN protein"					
FT	CDS	/note= "CDS does not include start codon"					
FT		complement (60238..61296)					
FT		/*tag= g					
FT		/*product= "NysD2 complete protein"					
FT	CDS	120628..121308					
FT		/*tag= h					
FT		/*product= "NysR4 (long) protein"					
XX	MO200159126-A2.						
XX	16-AUG-2001.						
XX	PD						
XX	PF	08-FEB-2001; 2001MO-GB00509.					
XX	PR	08-FEB-2000; 2000GB-0002840.					
XX	PR	10-APR-2000; 2000GB-0008786.					
XX	PR	14-APR-2000; 2000GB-0009387.					
XX							
PA	(UYNO-)	UNIV NORGES TEKNISK NATURVITENSKAPLIGE.					
PA	(SMTF)	SINTEF STIEPELSEN IND TEK FORSK.					
PA	(ALPH-)	ALPHARMA AS.					

PA (BRAU/) BRAUTASET T.
PA (STROM/) STROM A R.
PI Zorchev SB, Sekurova ON, Fjærvik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX WPI: 2001-557614/62.
DR P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR AAE10149, AAE10150.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
PS Claim 1; Page 188-254; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrocyclic antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
Query Match 43.6%; Score 44; DB 22; Length 125401;
Best Local Similarity 65.0%; Pred. No. 0.0018;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
OY 2 TCGACGTGCGACGAGGCTTCGCCGACGAGGCGCTCGACTCCCTGATGCTGTGGAGATCC 61
DB 109666 TCGGTGCGGGGCGACAGTTCAGAGTGGCTTCGACTCGCTGACCGCGGAGATTCC 109745
OY 62 GCAACGGCTTCAGGGTGAAGTATGCCCTGTCGGC 101
DB 109746 GCAACCGACTCAACGCGCCGACCGACTGCGCTCCGGC 109785
RESULT 8
AAZ87297
ID AAZ87297 standard; DNA; 13842 BP.
AC AAZ87297;
XX
XX 05-JUN-2000 (first entry)
DT
XX
XX S. venezuelae macrocyclic biosynthetic gene pikAI, SEQ ID NO:10.
XX
XX Desosamine biosynthesis; macrocyclic; polyketide; methymycin; pikromycin;
XX neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
XX chronic obstructive pulmonary disease; respiratory inflammation;
XX hypercholesterolemia; crop protection agent; ds.
XX
XX Streptomyces venezuelae ATCC15439.
OS
XX
XX Key Location/Qualifiers
XX CDS 1..13842
XX
XX /*tag= a
XX /product= "pikAI"
XX /transl_except= (pos:4156..4158, aa:Ala)
XX /transl_except= (pos:13741..14743, aa:Ala)
XX
XX MO200000620-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US14398.
XX
XX 26-JUN-1998; 98US-0105537.
XX
XX (MINU) UNIV MINNESOTA.
PA Sherman DH, Liu H, Xue Y, Zhao L;
PI
XX

DR WPI: 2000-160679/14.
DR P-PSDB: AAY77192.
XX
XX Desosamine and macrocyclic biosynthetic gene clusters, useful for, e.g.,
XX synthesis of methymycin and pikromycin -
XX
XX Claim 15; Page 377-383; 438pp; English.
PS
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrocyclic
XX biosynthetic gene cluster, or fragments thereof. The macrocyclic
XX biosynthetic gene cluster encodes proteins which synthesise methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
XX compounds. Recombinant or augmented cells comprising the desosamine
XX and/or macrocyclic biosynthetic gene clusters are useful for the
XX production of biologically active macrocyclics. The macrocyclic biosynthetic
XX proteins are useful for synthesis of methymycin, pikromycin,
XX neomethymycin and narbomycin. The alternative termination of polyketide
XX synthesis may be useful to prepare novel antibiotics and
XX polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
XX recombinant host cells are useful as biopolymers, e.g., in packaging or
XX biomedical applications, to engineer PHA monomer synthases or to prepare
XX biologically active agents, such as chemotherapeutics.
XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
XX disease as well as other diseases involving respiratory inflammation,
XX cholesterol-lowering agents or macrocyclic-based antibiotics which are
XX active against a variety of organisms, e.g., bacteria, including
XX multi-drug resistant pneumococci and other respiratory pathogens, as well
XX as viral parasitic pathogens, or as crop protection agents (e.g.,
XX fungicides or insecticides) via expression of polyketides in plants.
XX Sequences AAZ87295-287302 represent macrocyclic biosynthetic genes from
XX Streptomyces venezuelae ATCC 15439, which encode proteins
XX AAY77190-Y77197.
SQ Sequence 13842 BP; 1726 A; 5356 C; 4911 G; 1845 T; 4 other;
Query Match 42.0%; Score 42.4; DB 21; Length 13842;
Best Local Similarity 64.0%; Pred. No. 0.0038;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY 2 TCGACGTGCGACGAGGCTTCGCCGACGAGGCGCTCGACTCCCTGATGCTGTGGAGATCC 61
DB 13286 TCGACGCGGACGAGGCGCTTCGACGCTCGGCTTCGACCGCGGTGAGATCC 13345
OY 62 GCAACGGCTTCAGGGTGAAGTATGCCCTGTCGGC 101
DB 13346 GCAACCGCTCAACGCGCGCGGTGCGCTCCGGC 13385
RESULT 9
AAT80415
ID AAT80415 standard; DNA; 13987 BP.
AC AAT80415;
XX
XX 02-MAR-1998 (first entry)
DT
XX
XX Hybrid smg/cylG ORF1.
XX
XX Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
XX platenolide synthase gene cluster; platenolide production; smg gene;
XX polyketide; ty lactone synthesis; antibiotic; tylosin; hybrid gene; ss.
OS
XX
XX Streptomyces ambofaciens.
XX
XX Key Location/Qualifiers
XX CDS 350..13987
XX
XX /*tag= a
XX /transl_except= (pos:350..352, aa:Met)
XX
XX

/note="ORF1 encodes hybrid protein shown in AAW22611"

```
FT XX
PN EP791655-A2.
XX
PD 27-AUG-1997.
XX
PF 19-FEB-1997; 97EP-0301056.
XX
PR 22-FEB-1996; 96US-0012078.
XX
PA (ELIL ) LILLY & CO ELI.
XX
PI Dehoff BS, Kuhnscoss SA, Rosteck PR, Sutton KL;
PI WPI: 1997-418046/39.
DR P-PSDB; AAW22611.
XX
PT DNA encoding streptomycetes fradiiae tylactone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
PS Claim 22: Pages 178-197; 220pp; English.
XX
CC This sequence represents a hybrid gene of the invention. This sequence
CC was created by replacing a EcoRI-ApaI fragment of smg ORF1 with a
CC EcoRI-SfiI fragment from tylg ORF1. The position of the nucleotides from
CC each of the two genes is not given in the specification. The smg gene
CC (see AAT80414) was isolated from streptomycetes ambofaciens, and encodes
CC the multi-functional proteins which direct the synthesis of the
CC polyketide platenolide. Platenolide is the basic building block of the
CC macrolide antibiotic spiramycin. The tylg gene (see AAT80413) is the
CC tylactone synthase gene cluster of the invention. The tylg sequence was
CC isolated from streptomycetes fradiiae, and encodes multifunctional proteins
CC which direct the synthesis of the polyketide tylactone. Tylactone is the
CC basic building block of the antibiotic tylosin. The hybrid sequence can
CC be used to transform S. ambofaciens lacking the smg ORF1 sequence, or S.
CC fradiiae lacking the tylg ORF1 sequence, so that they can produce
CC polyketides. The DNA sequence can be modified so as to alter the type of
CC carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.
XX
SQ Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 other;
Query Match 42.0%; Score 42.4; DB 18; Length 13987;
Best Local Similarity 64.0%; Pred. No. 0.0038;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY 2 TCGACGTGCGAGCAGGCTTCGCCGACAGGAGGCTTCGACTCGATGCTGTGAGATCC 61
DB 13551 TCGAGGCCCGAAGCAGCCCTTCGCCGAGATCAGCGTCCGCTCGCGCGGTGAGACTGC 13610
OY 62 GCANAAGCGCTTCAGGGGAGCTGGGTATGTCGGTGTGCGG 101
DB 13611 GCNAACGGGCTCAGCCGTCTGCTGGGCTGTGCGGCTG 13650
RESULT 10
AAZ87318
ID AAZ87318 standard; DNA; 36778 BP.
XX
AC AAZ87318;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomycetes venezuelae ATCC15439.
```

```
XX
FH Key location/Qualifiers
FT CDS 1742..15583
FT /tag= a
FT /product= "Pik gene cluster protein #1 (AAV77200)"
FT CDS 15688..26907
FT /tag= b
FT /product= "Pik gene cluster protein #2 (AAV77201)"
FT CDS 26991..31679
FT /tag= c
FT /product= "Pik gene cluster protein #3 (AAV77202)"
FT CDS 31782..35822
FT /tag= d
FT /product= "Pik gene cluster protein #4 (AAV77203)"
FT CDS 35819..36664
FT /tag= e
FT /product= "Pik gene cluster protein #4 (AAV80997)"
XX
PN MO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MING ) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
PI WPI: 2000-160679/14.
DR P-PSDB; AAV77200, AAV77201, AAV77202, AAV77203, AAV80997.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e. g.
PT synthesis of methymycin and pikromycin -
XX
PS Disclosure; Figure 31: 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryc gene cluster of Saccharopolyspora erythraea or
CC Streptomycetes antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomycetes venezuelae ATCC
CC 15439, as given in figure 31.
XX
SQ Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;
Query Match 42.0%; Score 42.4; DB 21; Length 36778;
Best Local Similarity 64.0%; Pred. No. 0.0044;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY 2 TCGACGTGCGAGCAGGCTTCGCCGACAGGAGGCTTCGACTCGATGCTGTGAGATCC 61
```

Db 15027 TCGACGCCGACGAGGGGCTTCGACCTCGACCTCGACCTCCGACCGCGTGAACCTCC 15086
QY 62 GCAACGCGCTTCAGGCTGAGCTGGGTATGCCGCTGTCCGC 101
Db 15087 GCAACGCGCTTCAGCTCCGCCGGTGGCTCGCCCTCCCGGC 15126

RESULT 11

AA87285
ID AA87285 standard; DNA; 37948 BP.

XX
AC AA87285;

DT 05-JUN-2000 (first entry)

DE S. venezuelae plk (macrolide biosynthesis) gene cluster, SEQ ID NO:5.

XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;

KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;

KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;

KW chronic obstructive pulmonary disease; respiratory inflammation;

KW hypercholesterolemia; crop protection agent; ds.

OS Streptomyces venezuelae ATCC15439.

XX MO200000620-A2.

XX 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MIND) UNITV MINNESOTA.

PI Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI: 2000-160679/14.

XX P-PSDB; AAT77180.

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

XX synthesis of methymycin and pikromycin -

XX Claim 13; Page 299-315; 438pp; English.

XX The invention relates to an isolated and purified nucleic acid segment

XX comprising a desosamine biosynthetic gene cluster, a fragment or its

XX biologically active variant, where the nucleic acid sequence is not

XX derived from the ergC gene cluster of Saccharopolyspora erythraea or

XX Streptomyces antibiotics. The invention also relates to a macrolide

XX biosynthetic gene cluster, or fragments thereof. The macrolide

XX biosynthetic gene cluster encodes proteins which synthesize methymycin,

XX pikromycin, neomethymycin, narbomycin or a combination of these

XX compounds. Recombinant or augmented cells comprising the desosamine

XX and/or macrolide biosynthetic gene clusters are useful for the production

XX of biologically active macrolides. The macrolide biosynthetic proteins

XX are useful for synthesis of methymycin, pikromycin, neomethymycin and

XX narbomycin. The alternative termination of polyketide synthase may be

XX useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)

XX monomers. The compounds produced by the recombinant host cells are useful

XX as biopolymers, e.g., in packaging or biomedical applications, to

XX engineer PHA monomer synthases or to prepare biologically active agents,

XX such as chemotherapeutics, immunosuppressants, agents to treat asthma,

XX chronic obstructive pulmonary disease as well as other diseases involving

XX respiratory inflammation, cholesterol-lowering agents or macrolide-based

XX antibiotics which are active against a variety of organisms, e.g.,

XX bacteria, including multi-drug resistant pneumococci and other

XX respiratory pathogens, as well as viral parasitic pathogens, or as crop

XX protection agents (e.g., fungicides or insecticides) via expression of

XX polyketides in plants. The present sequence represents the macrolide

XX biosynthetic gene cluster (plk) from Streptomyces venezuelae ATCC 15439.

SQ Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

Query Match 42.0%; Score 42.4; DB 21; Length 37948;

Best Local Similarity 64.0%; Pred. No. 0.0044;

Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 TCGACGTGGCAGGAGGCTTCGACGACGAGGCGCTTCGACCTCCGATGTGAGATCC 61

Db 16197 TCGACGCCGACGAGGGGCTTCGACCTCGACCTCGGCTTCGACCTCCGACCGCGTGAACCTCC 16256

QY 62 GCAACGCGCTTCAGGCTGAGCTGGGTATGCCGCTGTCCGC 101

Db 16257 GCAACGCGCTTCAGCTCCGCCGGTGGCTCGCCCTCCCGGC 16296

RESULT 12

AA75633
ID AA75633 standard; DNA; 38506 BP.

XX
AC AA75633;

DT 22-JAN-2001 (first entry)

DE Nucleotide sequence of the insert DNA in cosmid pKOS023-27.

XX Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;

KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;

KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;

KW pikromycin biosynthesis; ss.

OS Streptomyces venezuelae.

XX US6117659-A.

XX 12-SEP-2000.

XX 27-MAY-1999; 99US-0320878.

XX 28-MAY-1998; 98US-0087080.

XX 22-SEP-1998; 98US-0100880.

XX 08-FEB-1999; 99US-0119139.

XX 20-MAY-1999; 99US-0134990.

XX 30-APR-1997; 97US-0846247.

XX 06-MAY-1998; 98US-0073538.

XX 28-AUG-1998; 98US-0141908.

XX (KOSA-) KOSAN BIOSCIENCES INC.

PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;

XX WPI: 2000-610844/58.

XX New recombinant pick hydroxylase gene of Streptomyces venezuelae useful

XX for converting ketolides to antibiotics and as antibiotics and

XX intermediates in the synthesis of compounds with pharmaceutical value

XX -

XX disclosure; Columns 15-32; 117pp; English.

XX The present sequence is used to produce the recombinant DNA compounds

XX expressing recombinant polyketide synthase genes in host cells for the

XX production of narbonolide, narbonolide derivatives and polyketides that

XX are useful as antibiotics and as intermediates in the synthesis of

XX compounds with pharmaceutical value. The DNA compounds may also encode

XX a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl

XX transferase enzymes (useful for conversion of ketolides to antibiotics),

XX and the beta-glucosidase enzyme (involved in pikromycin biosynthesis).

XX These compounds are also useful for increasing the antibiotic activity

XX of a compound relative to the unhydroxylated compound. The recombinant

XX host cells are useful as genetic systems that allow rapid engineering

XX of the narbonolide polyketide synthase. These would be valuable for

XX creating novel ketolide analogs for pharmaceutical applications.

```

XX SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;
Query Match 42.0%; Score 42.4; DB 21; Length 38506;
Best Local Similarity 64.0%; Pred. No. 0.0044;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 TCAGACGTGCGACGAGGCTTCCCGACGAGGCTTCGACTCCCTGATGCTGTGAGATCC 61
    ||||| ||| ||||| ||| ||| ||||| ||||| ||| ||| |||
DB 13169 TCGACGCCGAACGGGGCTTCTCGACCTCGGCTTCGACTCCCTGATGAGCCGCTCGAATCC 13228
    ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 GCAACGGCTTCAGGCTGAGCTGGGTAAGCCGCTGTGCGC 101
    ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13229 GCAACGGCTCACTCCGCGGTGGCCCGCCCTCCCGGC 13268

RESULT 13
AA256001 standard; DNA; 38506 BP.
XX AC AA256001;
XX DT 23-MAR-2000 (first entry)
XX DE Recombinant cosmid PKOS023-27 containing S. venezuelae PKS genes.
XX KM Narbonolide polyketide synthase; PKS; cosmid PKOS023-27; ketolide;
XX KW antibiotic production; narbomycin; picromycin; ds.
XX OS Streptomyces venezuelae.
FH FH Location/Qualifiers
FT FT /*tag- a
FT FT /product- PICAI
FT FT /note- "Narbonolide synthase subunit 1"
FT FT 13830..25049
FT FT CDS
FT FT /*tag- b
FT FT /product- PICAI1
FT FT /note- "Narbonolide synthase subunit 2"
FT FT 25133..29821
FT FT /*tag- c
FT FT /product- PICAI11
FT FT /note- "Narbonolide synthase subunit 3"
FT FT 29924..33964
FT FT /*tag- d
FT FT /product- PICAI1V
FT FT /note- "Narbonolide synthase subunit 4"
FT FT 33961..34806
FT FT /*tag- e
FT FT /product- PICB
FT FT /note- "Contains typeII thioesterase domain"
FT FT 34863..36011
FT FT /*tag- f
FT FT /product- PICII
FT FT /note- "4-keto-6-deoxyglucose isomerase"
FT FT 36159..37439
FT FT CDS
FT FT /*tag- g
FT FT /product- PICII1
FT FT /note- "Desosaminyl transferase"
FT FT 37529..38242
FT FT /*tag- h
FT FT /product- PICVI
FT FT /note- "3-amino dimethyltransferase"
XX XX
XX PN W09961599-A2.
XX XX
XX PD 02-DEC-1999.
XX XX
XX PF 27-MAY-1999; 99MO-US11814.
XX XX
XX PR 28-MAY-1998; 98US-0087080.
XX PR 28-AUG-1998; 98US-0141908.

```

```

PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX XX
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX XX
XX PI Ashley G, Belach MC, Belach M, McDaniel R, Tang L;
XX XX
XX DR WPI: 2000-072618/06.
XX DR P-PSDB: AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
XX DR AAY67208, AAY67211.
XX XX
XX PT New recombinant DNA encoding a domain of narbonolide polyketide
XX PT synthase, for production of ketolide antibiotics -
XX XX
XX PS Example 2; Page 16-27; 98pp; English.
XX XX
XX CC This is the recombinant cosmid PKOS023-27 DNA sequence which contains a
XX CC Streptomyces venezuelae DNA insert. The cosmid contains open reading
XX CC frames which encode the various modules of the narbonolide polyketide
XX CC synthase (PKS). The invention relates to recombinant DNA containing a
XX CC coding sequence for a narbonolide PKS. Polyketides are compounds
XX CC synthesised from 2-carbon units through a series of condensations and
XX CC subsequent modifications. Modular PKSs are responsible for the production
XX CC of many antibiotics including picromycin. The narbonolide PKS consists of
XX CC a loading module, six extender modules, and two thioester domains. Four
XX CC proteins make up the narbonolide PKS (PICAI, PICAI1, PICAI11 and PICAI1V).
XX CC PICAI includes the loading module and extender modules 1 and 2, PICAI1
XX CC includes extender modules 3 and 4, PICAI11 includes extender module 5 and
XX CC PICAI1V includes extender module 6 and a type II thioesterase domain. The
XX CC second type II thioesterase domain is found on the PICB protein. The
XX CC nucleotide sequences encoding all of these proteins can be isolated in
XX CC recombinant form from the recombinant cosmid PKOS023-27. Narbonolide is
XX CC desosaminylated in S. venezuelae to yield narbomycin, and the desosamine
XX CC transferase enzyme is required for this conversion, and the desosamine
XX CC biosynthetic genes are also found in cosmid PKOS023-27. The recombinant
XX CC DNA of the invention is used to express, in transformed cells,
XX CC narbonolide (or its derivatives) or other ketolides (particularly
XX CC hybrids), which may then be converted (e.g. by other enzymes
XX CC recombinantly expressed in the same hosts) to polyketide antibiotics or
XX CC their intermediates. The antibiotics are useful in human or veterinary
XX CC medicine.
XX XX
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;
Query Match 42.0%; Score 42.4; DB 21; Length 38506;
Best Local Similarity 64.0%; Pred. No. 0.0044;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 TCAGACGTGCGACGAGGCTTCCCGACGAGGCTTCGACTCCCTGATGCTGTGAGATCC 61
    ||||| ||| ||||| ||| ||| ||||| ||||| ||| ||| |||
DB 13169 TCGACGCCGAACGGGGCTTCTCGACCTCGGCTTCGACTCCCTGATGAGCCGCTCGAATCC 13228
    ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 GCAACGGCTTCAGGCTGAGCTGGGTAAGCCGCTGTGCGC 101
    ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13229 GCAACGGCTCACTCCGCGGTGGCCCGCCCTCCCGGC 13268

RESULT 14
AAT80413
ID AAT80413 standard; DNA; 43280 BP.
XX AC AAT80413;
XX XX
XX DT 27-FEB-1998 (first entry)
XX XX
XX DE Tyactone synthase gene cluster.
XX XX
XX KM Tyactone synthase gene cluster; tyIG gene; multifunctional protein;
XX KM polyketide; tyactone synthesis; antibiotic; tylosin; ss.
XX OS Streptomyces fradiae.
XX XX
FH FH Location/Qualifiers

```

FT	CDS	816..14243
FT		/*tag= a
FT		/transl_except= (pos: 816..818, aa: Met)
FT		/note= "ORF1 encodes protein shown in AAM22601"
FT	CDS	14351..19945
FT		/*tag= b
FT		/transl_except= (pos: 14351..14353, aa: Met)
FT		/note= "ORF2 encodes protein shown in AAM22602"
FT	CDS	20010..31199
FT		/*tag= c
FT		/transl_except= (pos: 20010..20012, aa: Met)
FT		/note= "ORF3 encodes protein shown in AAM22603"
FT	CDS	31232..36067
FT		/*tag= d
FT		/note= "ORF4 encodes protein shown in AAM22604"
FT		/note= "ORF5 encodes protein shown in AAM22605"
FT	CDS	36249..41774
FT		/*tag= e
FT		/note= "ORF5 encodes protein shown in AAM22605"
XX		
PN	EP791655-A2.	
XX		
PD	27-AUG-1997.	
XX		
PE	19-FEB-1997;	97EP-0301056.
XX		
PR	22-FEB-1996;	96US-0012078.
XX		
PA	(ELIL) LILLY & CO ELI.	
PI	Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;	
XX		
DR	WPI; 1997-418046/39.	
DR	P-PSDB; AAM22601-W22605.	
XX		
PT	DNA encoding Streptomyces fradiae tyactone synthase domain - for	
PT	production of tylosin-related polyketide compounds	
XX		
PS	Claim 2; Pages 8-66; 220pp; English.	
XX		
CC	This sequence represents the tyactone synthase gene cluster of the	
CC	invention. This sequence is also referred to as the tylg gene, and was	
CC	isolated from Streptomyces fradiae. This sequence encodes multifunctional	
CC	proteins which direct the synthesis of the polyketide tyactone, isolated	
CC	from Streptomyces fradiae. Tyactone is the basic building block of the	
CC	antibiotic tylosin. The DNA sequence can be modified so as to alter the	
CC	type of carboxylic acids incorporated, the number of carboxylic acids	
CC	incorporated and/or the post-condensation reactions performed, thereby	
CC	resulting in novel tylosin-related polyketides.	
XX		
SQ	Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;	
	Query Match	42.0%; Score 42.4; DB 18; Length 43280;
	Best Local Similarity	64.0%; Pred. No. 0.0045;
	Matches	64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY	2 TCGACGTGCGAGAGCTTCGCCGAGCAGGCGCTCGACTCCCTGTGAGTATCC 61	
Db	19512 TCGACGTGCGCGCGAGCTTCAAGGAGGGGGCTTGACTCCCTCAACGCCGTGAACTGC 19571	
OY	62 GCAAAAGCGCTTCAGGGTGAAGCTGGGTATGCCGCTGTGCGC 101	
Db	19572 GCACCGCGCTGCGCGCCGCCACCGGGGTGAACCTGCCGC 19611	
RESULT 15		
AA78508		
ID	AA78508 standard; DNA; 44377 BP.	
XX		
AC	AA78508;	
XX		
DT	26-FEB-1998 (first entry)	
XX		
DE	Platenolide synthase gene cluster.	

[illegible]

Wed Nov 6 09:54:30 2002

us-09-724-876-2_copy_20490_20590.rng

Page 12

Search completed: November 5, 2002, 12:37:51
Job time : 273.215 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds

(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_20490_20590

Perfect score: 101
Sequence: 1 ctgcagctgcagcagcgtt.....ctgggtatgcctgtcttgc 101Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

Genem1:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rtd:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF217189	AF217189 Sorangium
2	101	100.0	71989	6	ARI72664	ARI72664 Sequence
3	99.4	98.4	68750	1	AF210843	AF210843 Sorangium
4	99.4	98.4	68750	6	ARI93029	ARI93029 Sequence
5	99.4	98.4	68750	6	ARI99551	ARI99551 Sequence
6	99.4	98.4	68750	6	ARI99559	ARI99559 Sequence
7	99.4	98.4	68750	6	ARI99567	ARI99567 Sequence
8	99.4	98.4	68750	6	AR201097	AR201097 Sequence
9	99.4	98.4	68750	6	AR208671	AR208671 Sequence
10	77	76.2	10910	1	AX024383	AX024383 Sequence
11	77	76.2	10910	6	AX024276	AX024276 Sequence
12	52	51.5	20394	1	SNR133222	SNR133222 Streptomy
13	52	51.5	20394	6	AX067896	AX067896 Sequence
14	52	51.5	84985	1	SNR278573	SNR278573 Streptomy
15	47.2	46.7	41097	1	AF016585	AF016585 Streptomy
16	44	43.6	30690	6	E38020	E38020 Avertectin
17	44	43.6	64957	1	AB032367	AB032367 Streptomy
18	44	43.6	65140	6	AX211705	AX211705 Sequence
19	44	43.6	104526	1	AB070940	AB070940 Streptomy
20	44	43.6	113193	1	AF357202	AF357202 Streptomy
21	44	43.6	125401	6	AF263912	AF263912 Streptomy
22	44	43.6	125401	6	AX211739	AX211739 Sequence
23	42.8	42.4	78210	1	AB070949	AB070949 Streptomy
24	42.4	42.0	37948	1	AF079138	AF079138 Streptomy
25	42.4	42.0	43280	1	SEU78289	SEU78289 Streptomy
26	42.4	42.0	82746	1	AF453501	AF453501 Actinosyn
27	41.4	41.0	25150	1	AB070945	AB070945 Streptomy
28	40.8	40.4	30000	6	AX250263	AX250263 Sequence
29	40.4	40.0	104326	1	AB070940	AB070940 Streptomy
30	39.2	38.8	1681	6	ARI73328	ARI73328 Sequence
31	39.2	38.8	4015	1	AX024421	AX024421 Sequence
32	39.2	38.8	4015	1	AX024422	AX024422 Sequence
33	39.2	38.8	4015	6	AX024314	AX024314 Sequence
34	39.2	38.8	4015	6	AX024315	AX024315 Sequence
35	39.2	38.8	17101	1	AB017641	AB017641 Micromono
36	39.2	38.8	20235	1	SEERYAB	SEERYAB
37	39.2	38.8	20235	6	AR049368	AR049368 Sequence
38	39.2	38.8	20235	6	AR095529	AR095529 Sequence
39	39.2	38.8	20444	1	SEERYAB	SEERYAB
40	39.2	38.8	49736	1	AF319998	AF319998 Stigmatel
41	38.8	38.4	36181	1	U00023	U00023 Mycobacte
42	38.8	38.4	308950	1	MLEPPN9	MLEPPN9
43	38.6	38.2	59816	1	AF333038	AF333038 Streptomy
44	38.2	37.8	2237	6	AX330527	AX330527 Sequence
45	38.2	37.8	2237	9	S80437	S80437 fatty acid

ALIGNMENTS

RESULT 1
AF217189
LOCUS
DEFINITION
AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189.1 GI:7453554
SOURCE
Polyangium cellulosum.
ORGANISM
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 58733)

Pred. No. is the number of results predicted by chance to have a

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and
Julien, B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58723)
AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.
TITLE Isolation and characterization of the epothilone biosynthetic gene
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58723)
AUTHORS Julien, B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kusan Biosciences, Inc., 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
source Location/Qualifiers
1. 58733
/organism="Polyanjum celluloseum"
/strain="SMP4"
/db_xref="taxon:55"
<1.992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AA663878.1"
/db_xref="GI:7453555"
/translation="VGIQVAFADPGVPPVLLYDNLKNAVYRHGDAIPEHTIATL
SADYREPRVAVARGNIKGRVRAIRYRGFEFEARATYADLGLDKQATRETSAL
DRWEDKRAITQAFDHERSVLRHPDTPDHERVEVGTPTARELDNRYVPH
DTRRTLVLAIDISQVR.ADGNQIVAHVNSWDGQOIEPEHLQRLVDEKRAHRH
GIDRLAARSSQAFRLI.VAERGDNGSA.IARLIQDLIDVGAALAEALVEALERTI
HIGAVQVIDRRSERHIPPVSIPTVGRHAALVYTPHSLTTYDALKKDPT"
989.1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AA663879.1"
/db_xref="GI:7453556"
/translation="MTDITPTETKRLKSLGLPGLACWQLADKKPMLREVLAIEERE
RHKRSLERLKNSRVAIKMTDPSMWPKIDREAVDLDYRSYADLFEVYTRRD
AOKPLISTNKAFADMGQVPRHACVYTLVDRLVHRAVEYIEIAESYRLKEALNAT
RTKORRTKKH"
1998.6263
/gene="epoA"
/gene="epoA"
1998.6263
/note="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AA662880.1"
/db_xref="GI:7453517"
/translation="MADRIIERAAEDPIAIVGACRLPGVIDLSGFMTLLGESSRDTV
GGVPAERMDAAAFDPDLNAPGKTPYTRASFLEIDVACFDSPFGISIPREALRMDAHR
LLEVECMALLENALIAPIALVGTETGTCIGSEYEAALPRATASAEIDAHGGLTAM
PSVAGRTIYVIGLGRCAVADFAVSSSLVAHACOSLRSGCSRALGCVSLAMISPS
STLVMLSKTRALATDGRCAFEAENADGFRGECACATVYKRLSGARADDRILAVIR
SATNHGASGLTPNGSIOEIVLKRALADAGCASSVGIYEAHGTTIGDTEIDQ
LNAVGLGRDVATPLLIG:VKTNLGHPEYASGITGLKVVLSLQHQIPAHHLQAQLN
PRISMGDLRTYTRARTPIPDWNTPRAGVSPGMSGTNAHVLEAPATCTPPAPE
RPAELHLEARTAAALDAI.AARLRLHLETPYSCGLGDAVSLATRSAMENRLAVAT
SEBGLRAALDAAGQOTPI.GVVRGIDASSRGLAFLPTGGAQTILGMGGLVWPAP
REAFDLCVRLFNOELDRPI.REVVMMAEPASVDAALDLPATFOTPALFTPEVALAALRS
MGVEPELVAGHSTIGELVATCVAGVPSLEDAVPLVARGRLMOLAPGGMVSLTAPEA
DVAAAVAPRHASYSINAIV.GPDQVYIAGACQPHATIAAAMAAGARTKALHSHAFHS
PLMAVLEAFGRVAESVSTRRSIVLVNSDSCAGTDESSPEGVYRAHAREVREIDG
VVALHAAGTVEVGPCKSTLLGLVPACLPDAHPALASSRAGRDPAVLEALGGLM

ANGCLYSAGCLPESGGRVLPPTYPMQRRRYITDKADDAAGDRACAGADEYENK
GAVRGDRRSARLDHPPEPSGRREKVEAAGDFRLIDEIPVLDLTVLTVERRAPG
LGEPIALVADPAKLSFNDVOLALGMVDPDLPGKPNPDLIGGCACGIVLVEAGVGLY
VGCPVIALSAGAFATVTTSAALVLEPPALATNEAAAPVAVILAWYALMDLIDARLP
GERVLIHAATGGVGLAAYOMAGHVEAVHATAGTPEKRAYLESIGVYSDRSREY
ADVRAWTGGGVVYVNSLSGELIDKSFNLSHGRELGRDCAVADQGLDLPFLR
NLSFSLVDLRGMLEERAPVRLFEELICLLIAGVETTPPIATIPAVYADAFRSMAQ
AOLTKCTGLIDGPVOIRLPHAGSGPSTGDRDLDRASAKAPARAAALAFLEFTQ
VSOVLRPEIKGAELFTRLGDSLMARLNRTERASIKLSTLSTFSTSPRIALTLT
QNLDAIATRLSLERVAENLRAGVSDVSSGADQDWETIAL"
6260.10492
/gene="epoB"
/gene="epoB"
6260.10492
/note="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AA662881.1"
/translation="MTINQLNLEHOGVKLAADGRLOIOPAKNALPNILARISEH
KSTILMLRORLPASIVPAPAEHVPPLTIDIOSYMTGRTGATVPGSHIAREYD
CTDLDVARSRAFRKYVARHDLRAHTLPDMQVIERKVDALIEIIDRLGLSTIFRDL
RLVSLKRDASHRIYDTERPPLVHAVARLDEQOTRLVSLIDILNDLGSLSITFDML
SFYDEPSTLPVLELSYRDYVLALESRRKSEAHQRMWDYKRRVAVELPPPLPMKAD
PSYLRIRFRHTEOMLPSDSWSRLKAOVERGLTPGVILAAFEVIGRMSASPRETL
NITLENRLVHPHVNIDITGDEFTSMVLLDIDTTRDSEFORAKRIOLEAMADHDVS
GIEVOREARVILGIORGALFVPLVTSALNOQVAVYSLOLRDLPYVTSOTQOLLDH
OLEHNGDLIAMDIVDGVEPDDLDMDLEAVAFARLTREPSRMSQSLPQOLE
ARASAEPTLSISEHTLHGLFARVQDLNOLAVYARLTREPSRMSQSLPQOLE
BOGARNTIYAVYMERGMQVAVI.VALESAAVYPIDDIDABERTHILDLGEXKLY
LTPQWIDGLKSPGQIRLLVSDAGVEGALDPMPIPTDYSVDFVGLAAGGIIVVDA
GVWIDHGAVENTLIDINTERFIEGPGGVSILSLSDLEVTYVFGIAVGTIIVPDA
SKLRDPAHNAALIEREKVYVNSPALMGLVHESGRPSLARSRLSLTSGMDIPV
GLPELQALARPQVYSLGATGASLMSIGYPRVNDPSMAGIIPYGRPLRNDTPFLD
EALPEPPVWPGOLYIGGVGLALGYWRDEKRNSTFLVPEETGERLYKGDGRIYLD
GNIEEMREDNOIKIRGYRVEIGEIEETLSKPNVADIVIVPGDAANKILLAAVVP
EGRRRAAEDASIKTERVYDARNAKACGLSDGEPGRLCSLSVEPBGALPKRYISAG
TGLVPRACLDIVYARRRSVRLTEAPIPEVEGRFLCSLSVEPBGALPKRYISAG
STYVOTYVAVASGRLEGVDEGTYIHPPEHDLVSDGIERGAHVPONFVPEAA
FGLLEVGRIIDALIESLGSLSRECLLEAGVMMQMLEQAPSCNICVPGBOHDFDQVR
PVULDRHSDVYVHMGSRVDRPOFQVCTACVLMQDLSRRATTPGACRPOHDFDQVR
FLTKTLEPMVTFEVELDLPALTLVNGKDRALRERKOTSSPRSHGTAHPADALEI
LVAVVEPEVLEGVAGQSEFVDIGATSIHVHRSILQRLDREIATITELFCYPIGLS
LASGLRDKDKLQSEFRNMQDVRBARKKRRRS"
10639.16137
/gene="epoC"
/gene="epoC"
10639.16137
/note="type I polyketide synthase; epothilone PKS module
2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AA662882.1"
/db_xref="GI:7453559"
/translation="MEEOSSAIAVIGMSGRPGARLDLEFRNLRDGTAEVORESEQ
ELAASGVDPALVLDPIYVYRAGVLEVDVDFDAFGISIREBELMDPQRIPEBCAME
ALEMAGYDPTAEGSIGVYAGANMSYLLTSLNHEPRAMNRMGQOTLIGNDKDLAT
HYVRLNLTGPSISVOTACSTSLVNAHGLACMSLIDRECMALAGGTYVIRPARVYV
AAGGIESPGHGRAPDARKANGTIMGCGCVLLKPLDRLTSGDGPVAVITGSAIND
GARKICTFTRPSEYCOAIMEALMEALAGVARSIQYITERTGTTLLGDIAETAAARV
DKRSTRKACISGYKGTIGHLESAGINGLIKTVLALHROLPSPLSNEPSNEIDF
AASPEYVNTSLMDWNTSTPRAGVSSFGISGTNAHVLEERPAKLPAAAPARSAL
FVVSASALMDLPAANRLDHLQAHGSLSGVANSLATRRSPMEHRLAAMAAPAREAL
RGDLDAAGQCPGAVRGRCSPGNVKVFVEPQSGSOWVMGMLDAEEVFEALAL
SCDRAIOEAQMSLAEAECSQIRKIVDIVOVPUFALVAAALARSAGVGPDPV
VIGHSMEVAAAHVAGALSLEPAVITIGRSLINKKYGCGCAKVTLSASQNDALR
GYEDRSVAVNSPSTVYSGRPALITGVLSLANKVYSGCAKVTLSASQNDALR
EDLILALGCLPRRAAAYPRRSYVTGAMVAGPGLGANYMNNIRQYVRRAEYQADLOG
HGGLFENKSPHILITTSVEEMRAAQRGAAGVSLRQGDSEPAVLEALGALMAGQY
VPMGRLEFPAGGRVPLPTYPMQRRERYEAPKSAAGDRGVAGGHPHLEGMOTLST

QSTRMETTLDLRLPWLGDHVGAVFPGAATLLENASSGAEALDGPLOTIDVV
LAEALFAGDAVLYOVYTTTEQPSGRLOFOLASRPGHASFVHARGALLRVRTE
VPAGITLISVAPRLOASMPAAATYAEITFEMLOJGSPAGPAQIAELMRGEPALGRVLE
DAGSAAEYRLHPLLDLACPOVVGSLFAGGEEAPWVWVSCSELLORPSCGELMCHA
RVNHGROTPROGADENVVSSGAVVAEVSGLAORLPDGVRRREEDDWLEMEP
AAVGAARNAARMLLJGGGGGALRSMLEAGAAVHAESNTSAAGVRLAKAF
DGOAPTAVHLGSLDGGELDPGLGAGALDAPRADVSPALDPALEFARCGDVSMTY
QALAGMFRDAPRLMLTRGAAGVAVTQAPLLGLGRVIAEMHADLCARVDLT
PRLMVELTAAASRMLSLRYTAHRSAGAGDGDLLRLAARLAEARSGLLEPILR
AOISOVRLPGKIEVDAPLTSLGNSLMGLERIEAMGITYPATLLMTYPTVAA
LSGHARACEAPVESPHTADSVALEBESQDDLTOLIAKFRALTL"

gene
AF210843
16134..37907
/gene="epod"
16134..37907
/note="type I polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTGTPTAQQNPILKQAIIITQREIRLAGLQALETREPTAIV
GIGRFGGADAPAEFELDLDAERDAPDLDRMALVGAPEVAIPHAGLLEPTIDC
FDAEFGISPREASLDPQHLRLLEVAWEGEDLIPRSIDSGTVGVGAATLTERIDC
RTVARLPREEDAVSATGNMLSIAGARSLYLGLGQPCITVDTASLSLVAHLADRS
LRAGESDIALAGVSTLSPDMEEAARTQALSPGRCRTDASNGVRGCGCLVY
DKRTSDAQRDGRIMALIRGAINHDSRSTGLTAPNVIAOETVIAEALRSARHVEGAV
DYRTGCTGSDIPRIYEVALRATVGPARSCTGCVLGAVTNTIHLAASVAGLILK
AALSITHERITRNLFRLNRLRLESALALATEPVPTDRFRFAGVSSFGSGT
NAHVLEAPAVEIMPAPERSAEILLVLSGSEGLAQARLREHLMHPELIGDYL
AFSLATRSATMHLAVATRSREGLAALSVAOQOTPGAARCLASSSRKELLIFT
GOGAOTPGMGRGLCAAMPAREADRCVTLTDLRELDPLREMAEAESLILDOT
AFQTQPALEAVEYALTALMRSMGVPELLVGSIGLVAACVAGVSLDGVRIYARG
RLMGSLSGAMVSLGAEPAEVAANVAPHAMVSTAAVNGEBOVIYINBEAVQATIA
GFAARGVATKRLHVSNAHPSLMEPEERFVAVSVYRRPSVLSVNSLGRVYTD

Query Match
Best Local Similarity 100.0%; Score 101; DB 1; Length 58733;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGACGTCCGACGAGGCTCCCGACGAGGCTCCCTGATGCTGTGAGATC 60
|||||
DB 20490 CTCGACGTCCGACGAGGCTCCCGACGAGGCTCCCTGATGCTGTGAGATC 20549
|||||

QY 61 CGCAACGGCTTCAGGCTGAGCTGGGTATGCGGCTGTGCGC 101
|||||

DB 20550 CGCAACGGCTTCAGGCTGAGCTGGGTATGCGGCTGTGCGC 20590
|||||

RESULT 2
AR172664
LOCUS AR172664 71989 bp DNA linear FAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
Location/Qualifiers
1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 101; DB 6; Length 71989;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGACGTCCGACGAGGCTCCCGACGAGGCTCCCTGATGCTGTGAGATC 60
|||||
DB 20490 CTCGACGTCCGACGAGGCTCCCGACGAGGCTCCCTGATGCTGTGAGATC 20549
|||||

QY 61 CGCAACGGCTTCAGGCTGAGCTGGGTATGCGGCTGTGCGC 101
|||||

DB 20550 CGCAACGGCTTCAGGCTGAGCTGGGTATGCGGCTGTGCGC 20590
|||||

RESULT 3
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE
ORGANISM
Polyangium cellulosum.
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangiineae; Polyangiaceae; Polyangium.
1 (bases 1 to 68750)
Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milmanow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Stratemann,A., Cyr,D.D.,
Goriach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
Chem. Biol. 7 (2), 97-109 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSARVLDGHRATERHVLAERGIEDRLALREH
LRIOGGSFHCMDGLDITVELIADDOPLASISPHARSILRHPDMSAMVDDGPAIV
RWLARAGAPRLREYERERAKTRQERARMLAAAPCPADLPREDPDANGPLCP
MSPEVAEERKRSYATPELACALMLASVWSSKSGQGNIPALMEWELRTIVAMGADN
AIAASAPGTSSEALRGAARLFAESWVSSKSGQGNIPALMEWELRTIVAMGADN
LSREARAIAAEVRRLRAQAPFAAGAGLAVAGVSSGRLSGITVGDALVSGGND
IVMOPGRISIPVLAGTDPPELAPPLSOMLFVANAAGTISKVLTGDSPLIVARN
QARPMSTHVGCELSAGRLVEMRHHPHRGAASRRAYLSEHIIATWYSLTNATHVL
LACDEHLYWCELSAGRLVEMRHHPHRGAASRRAYLSEHIIATWYSLTNATHVL
WADPRRAITLVDRKTEVEPVLATRHPPAHVVSDDITAFITGQDPDSRQMHVEHTR
SGASTVADYORQDLMDRPDMLNRGLFTTNDRLTLARS"

CDS
1..68750
/note="ORF2; similar to hypothetical proteins from
Myxobacterium tuberculosis and Streptomyces coelicolor and
to DD-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"

```
/db_xref="GI:6724239"
/translacion="MGA:ISVAPGALGAEERBEGOGGODAGALAPAREVMAEVA
AGOMPAVWLVARGDDVINDVAVGTETLGSAPMBRDPRIETASMTXAVTVATVIMVE
EKKLDLSDPYDRMLPRELNRKVARIADIGPIDEIVPAPRPITTYRDIMFTNGRCSIPA
SSPIORAIIDELGLVNAQIVPMTPHGPDENIRKGTPLMHQPGQWMTNGSGISIPA
VGAADQGFDAFVREIRIAPLGRMDDEHVPADKLAFACGCFETDEQTEKTRMD
GASASAPSPASPGAA:LVSTVDYLLFARMINGVHGGRLLSKASVREKTDHL
TPAKKASFEFFGFETIWMGYMAVVTADPAVSEVGRYGMDCGFCTSWINDPGRRL
IGIVTOSAGFLFSGALIRFMRSVYVATESA"
3415..5556
/note="ORF3; simlar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAE26906.1"
/db_xref="GI:6724240"
/translacion="MHG:ITEROVLLSLVTALITVYARASGELARRLROPEVGLFC
GVYLGPSVYALAPGRH:ALFOEPANGVYLSGISGALLLLMAGIEVGLIRKEA
RQALSHLGAIPPLAA:AAFSALVDRPLPSGLISTIVAVSVIAVLIEREM
RRSYAOVTTLAAGVSEV:AVYLAMTSYSGASPALAVASALLSGFLFMVLVGR
LTHLARWVADATRVSK:QVSLVVLVFLAALTORGLHPLGAFAAGVLSAPRT
NRPLDGVOTLVAGLEF:VEFVLAMRVDSQLETPAAGTVALLLATATAAVYPA
LGARLGLRSEEAALVAGLNMKGTDLIVAYIGVELGLISNAYTMVAVALVYTA
SPALLIMLEKRAPPTOE:SALEBEEARAY:RGVRIILVPIYAHALCPADIVYTA
IVASKRIGETVITELSVQOAPGPRAGASRGLARIGARLVGTRRRELKRS
IQAILASRDHDLVIGANSFARAGSFEGLDQATVORAESVNLVVGDPAPERAS
ARRILVPIIGLEYSFAADLAHVALAMDEVLSSAQTDPGAIVVWRDREPRVRAV
ARSVDVAEVRGRGLGVHSSRVHVGHPDEITRELARAPYDLVLGCYDHGPIGLRL
YLGSITVESVVRVVALVHAGVREGVYR"
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAE26907.1"
/db_xref="GI:6724241"
/translacion="MDK:IGTRCAIAEGYIPGSGNPEPOMTSHEATCLNMSDRDA
OVAITYLESDRDPAGPYRTPARTRRHVFNDLTEPEPIPROTDASVLESDAPIY
OHRDLSROENMLSTIAYTORE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAE26308.1"
/db_xref="GI:6724242"
/translacion="MKHVJTGRRFRGRIHTLGLLASMALAGCGSPSEKTVOSTRLAP
GADARVADDPAAATV:AVDVVHLSPPRLKAGSREYVWQRPSPSESWRVAVLD
YNADSRGKLAETTVYAHFELLITAEKSSPSSAAVIGTFSVG"
7610..11875
/gene="epoa"
/gene="epoa"
/gene="epoa"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAE2619.1"
/db_xref="GI:6724253"
/translacion="MADR:TERAEDPIAIVASACRPGVTLDSGFMTLLGSRDTV
GVYAPARMMAAWEFDDP:ARGKTPTRASFSLDSVCPDASFGICISPRELRMDPRH
LULEVCWEALENAIAPSLVGETGVFTIGISEVEALPQATASIEDHAGGLCTM
PSYAGRISTALGRPCVAVDTAVSSSLVAHLLAQSLKSGCSFALGAGVLSLSP
STVLWLSKTRALARDGRCA:AFSAEDGFCGECACAAVVLKRLSGARADGRIILAVIRG
SAINHDGASGLTVPNSS:QEIYLRALADGCAASVGVYEAHGTITGDIETIOA
LNAVYGLGRDVAFTLLIG:VKTNLGHEVASGITGLKLVLSLOHGOITLHIAOLAOL
PRISMGDLRTVTRARTPK:PDWMTPRRAGSFRGMSGTNAHVLEBPATCTPPAPE
RPAELVLSARTSALDACAARLRDHLERYPSOCLDVAESLATITTSAMERLAVAT
SREGLRALDAAGOTSTGAVRSIADSSRGKLAFLTGGQATLDMGRKLDVMSAF
KEAFDLVRLFNQELDRPL:REVMMAPASVADALLDQTAFTQALFTFEVALAALRS
WGVEPELVAGHSIGELVAA:CVNAGFSLDEAVFLVAAARGLMQALPAGGAMVS:IEAEA
DVAAVAAPHAASVIAIVADQVYIAGAGOPVYHAI:AAAMAARGARTKLHVSHAFS
PLMAFMLAEGRVAESVYRRPSIVLVSNLSGAACIDEVSSPGYVWRHAREVYRFADG
```

```
VKALHAAGCTFEVEGPKSTLGLVPACMDARPALLASSRAGDEPATVLEALGLM
AVGLSVNAGLIPSGGRVRLPYPMORREYWTIDTADARDRAPAGHVEBEG
GAVRGDRRSARADHPPESGRREKXVAGVAGVATPEIDEPGLDHLVIRYMERAPG
LGEVEIYAVDAAGISFNDVOLALCMVEDDLPGRKNPPLLIGBEGAGIVAVGEBVNLV
VGPVPLALSAGAFRTHTTSAALVPRPOLSA:IEAAMPVATITAMVLAUDRIARLQ
GGERVLIJHAAGVGLAOWA:GVAGVAGVNAHTACTPCKRAYLESIGVAYVDSKDRFV
ADVNRATGGCGVGVVNLSSGELIDSFNLRSHGFEVEBELGKDCYADQGLRPLR
NLFSFVLDEKMLERPARALLEELILIAAGVFTPEPIATLPIARVADAFERSMAQ
AOHLGKVLTFDGPVOYIRIPTHAGAGPSTGDRDLDRLASAPAAAPAAALEAFERTO
VSGVTRPEIKVGAELFTPELGMDSIMAYVLEIRIASIKIKSTFTFLSTSPMIALLA
QNLIDALATNLISERVAENLRAGVONDEYSSADDDWEITAL"
7640..8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
11872..16104
/gene="epop"
11872..16104
/note="EPOS P; contains condensation/heterocyclization,
peptide carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAE26925.1"
/db_xref="GI:6724259"
/translacion="MTINOLNLEHQIKIADGERLQIOAPKALNPNLARISEH
KSVILMLROBPRESIVPAPARHAPPLTIDQESYWGSGAFVPGIHAYREYD
CMDDVPRLSRAKRYAVRDM:RAHTRDPMQVITEPKYDADIEITDLGDRSPBEA
RLVSLDANSRRTYDTERPPLIYVAVRDMERTRLVSLDINDLGSLSITFKRML
SFYEDDEPISLPLELSYRDVLALESRRKSEAHORMDYWKRIALPPLPUPMKAD
PSLTKIRIFRHRTEOMW:SDSWGRLKRVBERGLTPGVILAAFEVIGWMSAPRETL
NITLFNRLPVHPRVNDITGDEFSMLVDIDTTRDKSEQRARLQIOLWEMADHCVS
GIEVQREAAVLIGIORGALFPPVLTALNOQVYVFLRRLTEEPMGOVCSLPPQLE
OLYEHGDLVLAWDIVGVFPPDLMDMEAYVFLRRLTEEPMGOVCSLPPQLE
ARASANAVALYIEMHTLGLFAARVOLMOALVYVARRTIYEELSRSRRLGATLR
EGCARNTIYAVYMEKGEQYAVYLALESQAAYVIDIDPAPRERHTYLLDNGEYKLV
LTPWLDGKLSMP:PGIORLVSAVEGEGDDPPMP:IGTPSDLAIVITYLTSSTGLPK
GVWIDHRGAVENTLIDINEREIGDGRVLAHLSLSEDSLYDVDFGLAAGGIVPDA
SKLRDPAHMAELIEREKVYVWNSVPALMMLVEHFGSRDLSARLSLISLGDWTPV
GLPELOAIRPGVSVISLGATFASINSIGYPRANDLWASIPYGRPLRNOTFHVLD
EALERPWWVPGOLYIGVGLAGYWRDEKTRKSLVHPETGERLYKKGDDGRYLPD
GNTPEMGREDNOIKLRGYRVEIGEITEELTKSGENAVOYKRLAHGICRDLDKPYVDL
EGTRRRAAEDDSIKTERIDARHAAEADGLSDGENVOYKRLAHGICRDLDKPYVDL
TGDDPREAGLDVYARRRSVTRLEAP:IPVEFGRLSCISVPEGDATLPKFRYPBAG
STYPVOYIYVNSGRLEGVEGYHHPHRLKLSHDGIERGAVRQNFVFEAA
FNLLFQYRIDALESLYSGSSREFCLEAYEMQLMEQAPCSNIGCIPGCGNFEEVR
PVLDLRSIDVYVGMGLGRVDPFOVCTIGODSSRRATYTGAPGRGROHADMILR
FLRKLPEYWPVTFVEVELDALPITSNGYDRKALRRKOTSPSRBSGHTAPDADLEI
IYAVVREYGLIEVYGLQOSFVLDGATS:HIHVRSLLQKRLDREIATITFLQYPNLGS
LASGLRDSRDOLDORPNMDRYVRKRRRRS"
16251..21749
/gene="epob"
/gene="epob"
16251..21749
/gene="epob"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAE26920.1"
/db_xref="GI:6724254"
98.4%; Score 99.4; DB 1; Length 68750;
```

Query Match

```
Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 60
|||||
Db 26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 26161

OY 61 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 101
|||||
Db 26162 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 26202

RESULT 4
ARI93029 68750 bp DNA linear PAT 20-APR-2002
LOCUS Definition Sequence 1 from patent US 6346404.
ACCESSION ARI93029
VERSION ARI93029.1 GI:20238994
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 60
|||||
Db 26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 26161

OY 61 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 101
|||||
Db 26162 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 26202

RESULT 5
ARI99551 68750 bp DNA linear PAT 20-APR-2002
LOCUS Definition Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 60
|||||
Db 26162 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 26202

Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 60
|||||
Db 26162 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 26202

RESULT 6
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS Definition Sequence 1 from patent US 6355459.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 60
|||||
Db 26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 26161

OY 61 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 101
|||||
Db 26162 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 26202

RESULT 7
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS Definition Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"
ORIGIN
BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 60
|||||
Db 26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCCTCGACTCCCTGATGCGCTGGAGATC 26161

OY 61 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 101
|||||
Db 26162 CGCAACGGCTTCAGGGTGAGCTGGGTATGCGCGTGTGCGC 26202
```

RESULT 8
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 60
|||||
Db 26102 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 26161
|||||

Y 61 CGCAACGGCTTCAGGGTGAAGCTGGGTAGGCCGCTGTCCGC 101
|||||
Db 26162 CGCAACGGCTTCAGGGTGAAGCTGGGTAGGCCGCTGTCCGC 26202
|||||

RESULT 9
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 2.9e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 1 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 60
|||||
Db 26102 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 26161
|||||

Y 61 CGCAACGGCTTCAGGGTGAAGCTGGGTAGGCCGCTGTCCGC 101
|||||
Db 26162 CGCAACGGCTTCAGGGTGAAGCTGGGTAGGCCGCTGTCCGC 26202
|||||

RESULT 10
LOCUS AX024383 10910 bp DNA linear BCT 15-SEP-2000

DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024383
VERSION AX024383.1 GI:10184587
KEYWORDS
SOURCE Polyanium cellulosum.
ORGANISM Polyanium cellulosum.
REFERENCE 1 (bases 1 to 10910)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source location/Qualifiers
1..10910
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 1852 a 3124 c 3613 g 2321 t
ORIGIN

Query Match 76.2%; Score 77; DB 1; Length 10910;
Best Local Similarity 85.1%; Pred. No. 1.4e-09;
Matches 86; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Y 1 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 60
|||||
Db 2037 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 2096
|||||

Y 61 CGCAACGGCTTCAGGGTGAAGCTGGGTAGGCCGCTGTCCGC 101
|||||
Db 2097 CGTAACCGCTTCAGGCGAGCTGGGCGAAGGCGCTGTCCGC 2137
|||||

RESULT 11
LOCUS AX024276 10910 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024276
VERSION AX024276.1 GI:10184550
KEYWORDS
SOURCE Polyanium cellulosum.
ORGANISM Polyanium cellulosum.
REFERENCE 1 (bases 1 to 10910)
AUTHORS Beyer,S. and Mueller,R.J.
JOURNAL Patent: DE 19846493-A 13-APR-2000;
FEATURES
source location/Qualifiers
1..10910
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 1852 a 3124 c 3613 g 2321 t
ORIGIN

Query Match 76.2%; Score 77; DB 6; Length 10910;
Best Local Similarity 85.1%; Pred. No. 1.4e-09;
Matches 86; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Y 1 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 60
|||||
Db 2037 CTCGACGTGGCAGAGGCTTCGCCGAGCAGGCGCTGCACCTCCCTGATGGCTGTGGAGATC 2096
|||||

Y 61 CGCAACGGCTTCAGGGTGAAGCTGGGTAGGCCGCTGTCCGC 101
|||||
Db 2097 CGTAACCGCTTCAGGCGAGCTGGGCGAAGGCGCTGTCCGC 2137
|||||

RESULT 12
LOCUS SNA132222 20394 bp DNA linear BCT 24-APR-1999
DEFINITION Streptomyces natalensis pims1 gene.
ACCESSION SNA132222
VERSION AJ132222.1 GI:4678702

[illegible]

/product="hypothetical protein"
 /protein_id="CAC20917.1"
 /db_xref="GI:12055068"
 /translation="TRDAFREARETEPYGALLIAPILPERSADRVEGTKITSDMSQAREL
 AHDAACEVLLIAGQHPPLVLLIEDLQWADTAISDLRLYLSPROCGHGLIVLTARKRE
 TESIDLTRATVSDYLRGPRATLRLCGIPRAVALVEAHGSGVDEVEVYELVRSE
 GNPYVMOLLSTSYGDAVRHDPDAVDALHATIPGVEALHNMORPALPEEVLRLRCS
 AVIGAEVNTDLRRATGCDASVNGAGLESAIAGLIGEDPHPHQJLHAPVLLVNETTEL
 ELPRADRRRLIARVADALITARRRGVGDDEELERAHNHSMAKSLPAAYVLPRIILAA
 EBAEQMAIEVEYETILRRVAVHVGPIRPDDPATVLRQKHLIQIGQVLAATTRIGIHE
 AETATRRALSTVTHAEDPSVLMALCAASLVGRDASVAGSGDLSGLLRDLISWTRREV
 AVLGAAVEGIVLIVRGMLPOLALELIEGVEVGRFANEGRALTRQHPDPRVCRSY
 NTFETMILIGHQIDARRELSTLEYEERSDPSFALVYGVAAEGSTRTYLRLHL
 DEGVAVADDHGLIETWKAKMLGVYKGGKGVHADRDQEGALMHTSLAEIGSRTYLRHL
 HLGLIGQAHNNGQIEEARITTFRIILAAVEGRRRERYLHPRLATPTLIELHLLGNADE
 ALADDPERGCV"
 2190. .3566
 /gene="p1mk"
 2190. .3566
 /gene="p1mk"
 /function="putative glucosyl transferase"
 /codon_start=1
 /transl_table=1
 /product="p1mk protein"
 /protein_id="CAC20918.1"
 /db_xref="GI:12055069"
 /translation="MESARRPILFVSLPESGLNPLVLVLAGELSRQGEVDLMFTDEBP
 RNDKRLIAESPVPEAFISGVEDSEMSAVSSEVLEVTQPSYKKAHRAVARTYRPP
 QIOAKRFRRLIVDENOPALMVIDCISGFAGVDAAIRNPIYVSLPESNLTATL
 HFASKYTRGPPVPTHTGSRKQTLAQVRANLEFLRTFANPLNRLGKVALADENRRRN
 ELGPKASFMARIEHADLVLCNSLAEDYAPDITCEKRLVGAMVPLPEAPDDODLSR
 WLDASSVAVVYGLGITRLTREQVGSNEVAFARLEDHSHQVLMKLPSSQHLIPRESL
 PGNLESEVWSQMDVILAHNPVKVFTFHGGGANGNEGYTEGKPLVVRPLWVDCIDQAV

gene
CDS ..

complement(3662..9736)
/gene="pims4"
complement(3662..9736)
/gene="pims4"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="pims4 protein"
/protein_id="CAC20919.1"
/db_xref="GI:12053070"
/translation="MPNDKPNDKRLVDYLKWTWINDLHQTRQRLREVESAKOEPIATVGC
MSCRLPGGVRSPEEFWMLKLDKRDGIISFPPIRDWMDLIDLADGSSGSAAREGFTIDA
ASFDLFGFISPREAVTMDPOOLRLILLESFMALEPRAGIDPLVETGTFYGVGSGILG
YINTYMGSRREDIEGHSGLIASILSGSVYPTLGEQVAVDPAVDCSSLYSHLAAO
ALRVECSCLAAAGVTYATSTPMSPYSGROGLAPDRCARCAFTPTAGTAAGSECVLT
VVERISDNRKRGPIAVLRGSAVNODASNGLTALSGPSSQORTIROLAAAGIASYD
VDVHERGTGALDPILEQAQLATVATGDRDPEPRLPLIGSVKSLNHTQAAAGAAGLI
KMLILMRGPALERTLHPAEPSHVWDSSGALQLELEDEWPKADRPWRAGVSPAISGG
TNAHILLEGPPALPEEPEDEEPHTSPSPAPWLVASAKSEALATOLDRISALTQSPGLD
IGLSIATRAHLHERAVLVSEGGEVEVARAHAEHTAGKGLIYESGGOARVCGRGRL
YGRPEVPALELDGVLAHRTDGLRDMVEEALRDEGTGTQALPEALVAFALRESLJG
VRPDPVACHSIGELIAAHVAVCYFSLADACALVARRALMOLPRGAMVAVOAAESPV
LARLDVRSIAVNVGPEAVVATAGEBEYRLTAEFTGLGRTRRLPVSHPHSTLMP
MLDEDRRAVEGISTEAPRIATVASNLTGELATEDVRSKEVWRVRVETPRADSVSLP
LAEGSESTLELGGDGLVLTMAAOHTDGLATVATPALRKDSERTALRLLAOLAHVAGG
VDMSSIFPGTSGARVLDLPTYPEOHERWYPRPVAIVSGDVAAHBLTARTGLLOEMSG
ESGEVLELVGRLSLVAPMLTDLHTMGSGVLSFAPATGFLAMAAOVCDCRDREVDLTP
PLVLAESGTOLOVYGRPADAGRTTISIARPGDPDOPMHOAGGGLTISGEIVTQF
GSTVWPGCAENVDLSEFTYDTTGCGPAQCGIRSWQCEDDAYVVALPDVABDAGAF
GHPVLDLDAVLSRRMAGTGAAGDTLFFFAVRGSLHAGASVATLERNVIRKGDGTL
VAVDEGGAPLYLSREVELYRLQTLHTDPAFLARKARADALLSLEWVGPEPPAGVAC
VSGTADLEGVARSVATLADLSGDEDFVYVVGSGDVPAAHBLTARTGLLOEMSG
REGTSALDEVATVTRNAVADGSGEVRDPAALAAWGLVSAQSESPGRITALVDLEADSG
DLAEVLLPLGLIAGDOAFVYRGAVVGRLVESNEGAGCPREMDAGCVLTGGGT
GGLGELIARHLYVERGVPHLLVSRSGEAGSGVELEDELAHAQELAVACVADADA
AVEAVANGVDAHPLTAVITRAGVLDGCVIASLTPERSAVLRKYDVAANNLHEATKD
LDLAEVLYSSISVGLSGQGNAGVNFLEDALAGHAGTGLVAVGRLAMCAMPVGGP
MTATLSDMDQRMASYSTGAPLSPBERGLAFEPAAAGLVAHLYVAGRLSGVAPRAGVGGP

SHLRGLKGAARTAVNTTGETAVSLADRLRAARPEREOTQMDTVETEAAYVLGHASS
QNVADAREEYELGDSITVELRNRLSTVTLGRLSATVVEFDSKNPDLARLYADLAA
OSALDGSASQAGELPVAAPETDSLEBEMFLDSDGKVPKQRLSALGALRPFSEMTA
ELEDDLPTATLAEAGPGTPLRICVSTPTANGVHEYARLAAFPGERBHEVPLPVGAT
GERLPATTAETAVVIAESALRASDGNPVLVGHSSAGAYLAALAEELTVMGCRPAT
VLLDPTSLRHEONESTIYAGLARRHREKVDVSVRYNTRSLMAARKMGLNOLVRRH
TVVPLILTRATKETFGITGDTGTYGEDHSPVDVRSVDADHSEVMRDAPETARIVRE
WLDLSIGNA"
/complement(9772..15198)
/gene="p1ms3"
/complement(9772..15198)
/gene="p1ms3"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="p1ms3 protein"
/protein_id="CAC23920.1"
/db_xref="GI:12051071"
/translation="MTT3SETVVAALRASLKAEENLRONRKLVAATEPIATVAMAC
RPGGVQSPEDLWELTLAGDAIGAPTRGMKMDLRADGVRDNAAVSOEGFLDG
VGDFADPFGISPRESVTMDPOQLLETSMWEIERAGIDPTLRGRRTVGTNGQ
DYAYLTVRSIADATGDIITGIMASATSGRLSTYLGEGPAVTVDTACSSSLVLAHNS
HLRAGECSLAVGVWMSRPSGLMEPSRGGIADGCKAPADDCMGSEGVY
LYVERLSDRRNHNPVLAIVRGSANQDGSNGFTAPSGSOORVTRQALAAAGLSAS
DVDVVAHGIGTPLSDP..BAQALATVYGGDRDPERLLLGYSKSLGHOQAAGAAGL
IKMLAMRGVLPKTLVYDVPSSHDVMTGSEVLELPQMEPKADNPRQAGVSFEGIS
GTHAVIIVEQAEDEEPVAPLPAVAPVPMVSAKTEELAAQALARISSRGTAASVD
VGFSLASGAFERHRAVLAOCQDGLSEARARATSPRLATFSGOGAORVGMRE
LVYRPPVAPALDEVALHFDCAIDRVMDRDEBDLDTGTQPLALPIEVALPRAESE
GYRPDVGCHSITETIAAHVAGVSLTDCAVLAARAKMALPACGAANAQVASE
VVARLDVGSVIAAVNGPIAVVLADGETAVLRIAEELAAAGKRTQRLPVSHAFSPIME
LMEGASAFLELPGDQVLTAMAHHTLDEKVAIVPKRDEETALLTALHLHVAIG
LWMSGIFPAFTGARGVDLPYRQHERVMPRPVLSGVDAAGLRSABEHPLGATLTL
AESGEVTLGRLSIOTHIWLTQHAIGSNPEFTGTETELMAHAADACDVRDVIAM
TPIVLAENGTQVQVLCPADAGCTTITIHAREDDPQPMQAHGGLTUSERVYD
FGSTVWPQDDAAVADLEGEYEGTGYPAQGLSVSCDGAIEVALPDHVAEDGA
FLDPLALDLAVLOSRIAGVADQFLPFAWGVSLHAGASVLRVVKGTGDDNS
IAVADDEGAPVLSAEAMVRAYRDPCTQAAAHGAREDELSELEWSPPEVPACVR
CYSGLADELGVASVASLADSGEDFVVLVPSGVGDVPAAHAEHTARTLIGLOMS
GREGSGSLVFTVTRNAVSADGESVDRVAAAVMGVLAIVRSGESRGTALNDEMS
GDLAEVLAELPGILANGDROFYVREGAVYVGRVLESNEGAGSPREMDVGYLVYTG
TGLGRELRLHLVVERGVRLHLLVSRSGEPAPVSLDSLELAHGEVTVARCDVADR
AAVHALVAAPPEHPJLVAVIHTAGVLDLGLVSLPERSTVLRPVVDAMCHXATK
GDLDAFVYSSISGVLAGQOANYAAGVNFDLPAACRAMHMGGLPQOSLACMGATDG
GMTATLSDTDMOMASGYTAPLSRORGILFDPAACRAMHMGGLPQOSLACMGATDG
PSYIRGLVGARRTAVNTTGETAVSLADRLRAARPEREOTQMDTVETEAAYVLGHAS
AKNVADAREEYELGDSITVELRNRLSTVTLGRLSATVVEFDSKNPDLARLYADLAA
EPAATGPALLAELELRDSALAGAPDARTAAVTAALITDMDRDAPEDEGAEVAE
RLEAATDEIFAFIDNELJRHSDR"
/complement(15210..43733)
/gene="p1ms2"
/complement(15210..43733)
/gene="p1ms2"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="p1ms2 protein"
/protein_id="CAC20421.1"
/db_xref="GI:12054072"
/translation="MDQEKLDYKLRASADLRSSQVGELEAAREPIATVGMSCR
PFGGVSPBELMSLADADIGITKSPADGWEMLADAGEFSGFLHDAAEFADSPFG
ISPREALAMPQOURLLEISWEAFERAGIDPAARCTGTGMFGAMPQOYRVGPDSD
AGVALTGTTSYISGRLAYFGAVGPAVTVDTACSSSLVLAHLSHAGRECSLALA
AGVWMSPTTVEFERBOGGLSADGCRSPFADSCSTGTGSEGVYVLERLSEARNG
HEILAVIRGSAYNQDASHGLTAPNGPSORRTIEOALVSARLSADEVDVVEAHGCTT
LGDPEAQLATVYCOGRDDBRLPLTGVSKSLNGHQAAAGAVTKVLAARHGVLP
RTLHNDAPSSHDVMTQAGRLRLTEHLNPGQGBOPRRAGVSFGLSGTNGTILIEEPA
APSPGAGQEDTAPVAEPLAAGVPMVLSGRTREALRAQARLLEHLTSGPLADLPA
YSLATRSGLLEHRAFTAGPGARALALADGPAAGLVODTARTTRKLAFLFPGQ
GSGORGMRELAARPPVFAATADVLGHFDEGLERLKVLEPAESTPPAALLDQTY
AOPALPAIEVALYRLAESFGITPDLAGHSICEIAAHAGVFTLADAAALVLAERL
MOLPBGAMVSLATEEVEVLPDLLEDRPRLSIAAVNGPAAVIAGEADVTVAVERH

DALGRTKRLVSHAFSLPMEPMIAEFTAVSRLPQAPVPIVSGLTGLATVEQL
TSDVYLDNARHARVPADALRLARRRRHNLPRKMPDGLSAMASCCGCHGRRRR
LBNARVARGREAEETLTAALGLNARRRHNLPRKMPDGLSAMASCCGCHGRRRR
PSIPATGGDVNARAGGAHHPLTLAAVSAVSDGILLGRSLRTHMLADHAYRG
VLLPGTAFLELAVRAGDEAGCGRVEETLTAALVLPREGVOYVWMSPPASGRRAV
SVHSRDPGEEELPWTHAGTIVAGCHNAGFPATVWPPADATPLLDGFTYAMADTGF
GYGLPQGLRAAMRSGDDVYAEVALPDSGAASTSSAAGLHPALDLAAHAAAFV
LGEDARGGLPFSMODVTHASGASTVRYVLPDGDVAALVADPTGAPVASIGSLV
RTLPEQIDIAASHLVLDLFEELRMSARQTAHEAPASAVALPDLFGLERLALPDL
VTAVRGPDAAPAHAPDVLAIPVAGTDEDDVTGVNARTAHVLERLTHMLDERAASRL
VYTRGALATDARQASDPVAAAVMGVLAIVRSGESRGTALNDEMS
LALSDPEOTAVRGTAVTLARLARAQAPATATMPDEGVTVLTGATGGLGRVLAHLV
AERGVRLHLLVSRSGADADGAEELAAHGAQVYLAACDLADRAARRLISGVPAE
HPYTAIVHSGASLDDGVIALSLTPEPISAVLRKADAVMALHEATBDLDAAFVSSVL
SGTVAGAGOGANYAAAFEDALAOURLRAGGLGLSLAMGPAPATAGMTGLDDBDLR
LARMGTPALTEBOGIALFPAATVTOAAVLPTRIDLSYVRAVOSSEIPLMRGLIRTPAR
RSVYSASETVAITVLOLSRLATPREDREVLDIYAOVAAVIVGHACTADIDPPRREL
GFDLSLAVELRNRLSATGRLRTAAVYIFHPVVDALAAHLDLGLMSERAAAAADPRT

Query Match 51.5%; Score 52; DB 1; Length 84985;
Best Local Similarity 70.0%; Pred. No. 0.002;
Matches 70; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 2 TCGACGTGCGACGAGCGCTTCCGAGCAGGCGCTCGACTCCGATGCGTGTGACATCC 61
Db 76684 TCGATGGCCCAACGCGCTTCCGAGCAGGCGCTTCCGATGCGTGTGACATCC 76625
QY 62 GCAACGCGCTTCAAGGGGTGACGCTGATGCGGCTGCGC 101
Db 76624 GCAACGCGCTTCCGCGCGCAGCGGCGCTGCGGCTGCGC 76585

RESULT 15
AF016585
LOCUS
DEFINITION
Streptomyces caelestis cytochrome P-450 hydroxylase homol (nidi)
genes, partial cds; polyketide synthase modules 1 through 7 (nidi)
cds.
ACCESSION
AF016585
VERSION
AF016585.1 GI:2558836
KEYWORDS
Streptomyces caelestis.
SOURCE
Actinomyces caelestis
ORGANISM
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1 (bases 1 to 41097)
Kakavas,S.J., Katz,L. and Stassi,D.
Identification and characterization of the niddamycin polyketide
synthase genes from Streptomyces caelestis
J. Bacteriol. 179 (23), 7515-7522 (1997)
AUTHORS
Kakavas,S.J. and Stassi,D.
TITLE
Direct Submission
JOURNAL
Submitted (30-JUL-1997) 47P, Abbott Laboratories, 100 Abbott Park
Road, Abbott Park, IL 60064, USA
FEATURES
source
1..41097
/organism="Streptomyces caelestis"
/strain="NRRL-2821"
/db_xref="taxon:36816"
complement(<1..396)
/gene="nidi1"
complement(<1..396)
/note="orf6; similar to ty11"
/codon_start=1
/transl_table=11
/product="cytochrome P-450 hydroxylase homology"
/protein_id="AAC46023.1"
/db_xref="GI:2558837"

gene
CDS

```
/translation="WVDSVTGPMELSKDNANKELLDWFSHNRTHIIPVWDEGRQAMV  
FRYDYLTVSNHNPFFSSDFTEVAPTPELELITLGGPTIGALDPAHPGMKRLVSGAF  
TPRRAGGEORITVIAEELLDRVGRQKITA"  
627..113649  
/gene="n1da1"  
627..113649  
/gene="n1da1"  
/note="orf1"  
/codon_start=1  
/transl_table=11  
/product="polyketide synthase modules 1 and 2"  
/protein_id="AAC46024.1"  
/db_xref="GI:2558838"  
/translation="MAGHDATPAOKAODAEKSEDSDAIAVIGMSCRPPAGTAPEF  
OLLSGADAVYTAADRRRCITIDAPADPAAPFGMSPREMAATPDQQLVLELGEAL  
EDAGTVPESLKGEASVFGAMNDYATLLHRAFPDTYATATGLOSHMINRISYFL  
GLRGSILVVDVGQSSSLVAVALAEVSLRGSTGSLAGAVMLVLAEESSAMERAGL  
SPDRCRHTFDARANGYVREGGAIIVLKPPLADALADGRVVCVVGATGNDGGPGL  
TVPDRAGEAVLRAACDQAGVPADVPVELHGTCTPGDPEAEALGAVYGTGRPAN  
EPLVGSYKTNIGHGGAAGIAGFYKALCLHERALPASLNEPFPNAPILPERLKY  
QTAHALQPGTGGPPLAGVASAFGMGCNCHYVLEPTGCRQAPATGQADACIFASP  
MULISAREQALRQAARLRHLEDSDGADPLDIATSLATTRFERHRAVPGGDDRL  
SSALAALAGQTPRGVIRGSTDDGRALLLTGQAQHPGMQELVTTDHPHAALE  
VCEELRCGTQNLREVMTPDQPLDRTETYPALFALQIALYTLTARQTOALVL  
GHSVEITAAHIAAGVLDLPDAARLITTAHAWGOLPHGAMLSVOAEEHDDLOLHTH  
GVEITAVNGPTHVCSGPRALAEETAOHLREONVHTWLKVSHAFSALMDPMIGAFR  
DLNTIANVQPTPIPLISNLTCQIADPNHLCITPDYITDARHTVRFADAVOTAHQGT  
TYLEIGHPTITLLHHTLDNFTTIPTLHREPEPELTQATAAGVTDGIDMAVLC  
GASRRVRELPTYAQRTHMAPGLTPHNAAPDRAAEPRAMAGVPSRELVNLV  
ETIASVLGDPEVALDRPTSGLDSTAVELAGLITAGVALDPLTYLEPTPR  
AVADHIAKTLIGESADADQENGRTEGAEKAGDPIAVIGIGRFGCVATPDOLME  
LVASGTDAISTEPDGMWDGLVDPDSTPGKSVVRHGGFLHNAOPDAEEFGISPR  
EATAMDPOORILLTSMWALERAGVPSLIRGRTGVFTTAPERYGRHLHGTCYE  
GFLTGTTASVSGRIAYALGTRGAPALVDTRKAFSADASTAAEGGMLVERLSDAEULAG  
GTVMSGPMFVEFSRQGLAPDGRKAFSADASTAAEGGMLVERLSDAEULAG  
RVLAVVRGTANODASNGLTAPSGPAQOOVIRALSDAGSANDIDIVERHGTAL  
GDPIEAGALATYGHPRKOTPYMLGSLNSIGHQQAAGIAGIKMYOALRHDTLPT  
LHADHPSKVMDAGPQLLTDARPMWADPDRPRRAGISAFVSGTNVAHVEEPPAP  
BEAPRQAPPPVIAWPLSAHPTPALRQAARLDSMLHHTQADVODIAHALATYTHF  
KHRAVITRTRELTKHTIDALDQGTAHPPRLTLTGGAQHRGMOELIATDPH  
FAALDEVCEELQRCQNLREVMTPDQPLDRTETYPALFALQIALYTLTARQTOALH  
TOAHLYLGHVGEITAAHIAAGVLDLPDAARLITTAHAWGOLPHGAMLSVOAEEHDD  
LOLHTHGEVETIAAVNGPTHVCSGPRALAEETAOHLHQOQIRHFWLKVSHAFSALMD  
PMIGAFRDTLNTANVQPTPIPLISNLTCQIADPNHLCITPDYITDARHTVRFADAVOT  
AHQGTTYLEIGHPTITLLHHTLDNFTTIPTLHREPEPELTQATAAGVTDGIDMAVLC  
TYLEIGHPTITLLHHTLDNFTTIPTLHREPEPELTQATAAGVTDGIDMAVLC  
TTLHRTSPQTHHDLPTYPFORDRIMEPVRAYOVSQGPAGDRMLRYVWEAPARDHT  
VPSARMLLDGSDPADAGILVGEEOVLVGAOVRDLIPAPGRILLSCSELTPVR  
GSHTEGFCGRSLRSPRYAEVALLQALDAGADAPLWATRGVAQAQAGAPVSGA  
OLMGIGVYAGLELADRMGLVLDLPADPAARLGLARVLASNAQNOYAIRASGVFR  
RVPAVGRVNDNAPSGTVLITGGTALGSOVARRLAGAPHLLLAGRGNSGAE  
LVDELTLGAETVYACDAADRDAIVSLATIPHRPLTVYIHAAGVLDGCVLDTLP  
ERIDAVYBAKATAARHDELTDLDLDFVFSYLVGYMNGCATYTAANALDADALA  
HGRRAQGRATSIAMGPWAGSGMAAGDAKSLARGLDYLVEEALDLLEAVRDET  
TLVADVDMETLSSASRRQLPFGRIPARAVAGADGSPDVVWFHSPQSLAA  
EROLRLDLVRETAASYLRHGSATVDAETAFRAGFDSLITVELRNLRLASGTGLRS  
NTLEDDHPTPAVAEYVLEOILGTEPSAASGAGARTVVASDEPAVVAAGCAVPG  
GVATPDDIMELVASTDAISTEPDGMWDGLVDPDSTPGKSVVRHGGFLHNAOPDA  
DAEFGISPREATAMDPOORILLTSMWALERAGVPSLIRGRTGVFTTAPERYGRHL  
QSRKADYTGHLITGLTASVSGRIAYALGTRGAPALVDTRKAFSADASTAAEGGML  
RGECDLALAGSTVASTPTVFEFSRQGLAPDGRKAFSADASTAAEGGMLVERLS  
RLSDAERLGHVLAIVRGTAVNDGASNGLTAPSGPAQOOVIRALSDAGSANDIDIV  
VEAHGTGTPLDGPTAEALATYGHPRKOTPYMLGSLNSIGHQQAAGIAGIKMYO  
AMHNGTIPRTLHVAPASQVDMWEGSLOLAEAPWADPDRPRRAGISAFVSGTNVA  
HVLVEAPRPAEPARQAPPPVIAWPLSAHPTPALRQAARLDSMLHHTQADVODIA  
HALATYTHFRAVITGRTRTELTKHTIDALDQGTAHPPRLTLTGGAQHRGMOELI  
ALYRTLTAHQTOALVHLSVGEITAAHIAAGVLDLPDAARLITTAHAWGOLPHGAM  
LSVOAEEHDDLOLHTHGEVETIAAVNGPTHVCSGPRALAEETAOHLREONVHTWL  
KVSHAFSALMDPMIGAFRDTLNTANVQPTPIPLISNLTCQIADPNHLCITPDYITDAR  
HTVRFADAVOTAHQGTTYLEIGHPTITLLHHTLDNFTTIPTLHREPEPELTQATA  
LATLHTTGHTTTPHPSHIAPOVRSLPAPFORRAVMPNSAAHIGRSDAEATRTGLA  
RODHFILTGATSIYSGATLLTGRVSLAHPMLADHTVAGAVLFGTAPADLLIVAE  
ETSGTVEELTLHWPLVLPAGQAVOGLVLEAPDERGRRAATVYARPEDEETGAEWTH
```

gene
CDS

```
AAGVLAGAEPAAPAVDMADMAAGAMPPEPGAADVDELYAEFAAGYDYGPAFTGLT  
AMRGEVEFADVRLPEAVGAGEEERFRGPALEFADLHPMRAGVPMQESGPADGG  
OSTLPEFWOGUCLTYGAEQVRYNLAISDCAESVRADARAGVPLADLALVLRVP  
SSTPGALPAAPMTHIDOPVDSISTVOTRRVAVGADDETAEVIRATLPRAHTTAAH  
DLAALRAAPDAVAPPPGALVVALIPGSPAPPEPTSAVDPRATLGLSLVQALIE  
DERLTGTTLTAVVRGAEVATQTDVDPDLAGALMOLVSAQSEYGRRLIDVDVSE  
SRALATATLASEDPOLARAGVMLAPLAPLNPESALRELELTELGVHSFSCDLAD  
TGALGRVAVHILARRHGVRRMLVRSQPNPESALRELELTELGVHSFSCDLAD  
ADVRAVAGVPSDHPILGVYHTAGVLDGALARLAPRERIDTVRAKADAVRHLEART  
GHPLEAVYLSAAGLGRPGOASTAANAHVADAFARLRARAGVPASLANWHDERE  
GMAAGLDATALLRLRDGILPMPADQALDELDOALITTYRDRDPALLVPLLDLSTALRT  
AGENPASPVPPLRLSLVPSHPRHRTTATSVRDSGSPSPSARLALPAKRMVLDL  
TVSEOVAGVPLGHLSAELDEPPEPFIEIGFDSIALTELENNRLGRVGLPLSTLADHP  
TPRKAEMINIMELPANPALGSAEDPVLITGIDOWARAIALMDADGRADYRQVLG  
LITLADPRASGCDLRAAATTVADRLDADTDEITAFIDBQLEOPDPLCAPLDERL  
DEQP"
```

```
/codon_start=1  
/transl_table=11  
/product="polyketide synthase module 3"  
/protein_id="AAC46025.1"  
/db_xref="GI:2558839"
```

gene

CDS

```
13693..19212  
/gene="n1da2"  
13693..19212  
/gene="n1da2"  
/note="orf2"  
/codon_start=1  
/transl_table=11  
/product="polyketide synthase modules 4 and 5"  
/protein_id="AAC46026.1"  
/db_xref="GI:2558840"  
/translation="MSATNEELKREYLRAMCDLHSAERLAELSESAREPAAVYGM  
CRYPGCVATPDOLMELVASGTDILFTPTDRWDGLVDYDDSPPTGKSYRHGGFLH  
DAAQGDAAEFEGISPREATAMDPOORILLTSMWALERAGVPSLIRGRTGVFTTAP  
HDYGHQVGSAAADSGOGLGTAGVAGRAVAYTIGLGPALTYDTACSSSLVALHLA  
VQSLRGSCDLALAGGTTVMATPTVFEFSRQGLAPDGRKAFSADASTAAEGGML
```

MLVERLSDAERLGRVLAVERGTAVNODGASNGLTAPSGPAAQOQVIRDALADAHLP
DDIDVAEAGTGTPLGPIEAGALATYGHPRQTPVWLGSLKSNIGHTQAAGIAGI
IKWQALRHDTLPRLTADHSSKDWEGAPLQILAOARPPADDPRRRAGISAFGV
SGTNAHAIIEEPTLFEAPADTHPPVLTAMPILSAHTPTALRAQARLRARLHRTSDAL
SLADSAGTGHGLAVGRALPHRAVLIGDGAAPLDALALASGEYSPPVYVGSADVR
VAVFPGQAGWAGMG#ELIDSSVFNAELARCBALFVDMSTIDVLRGAPGCL
DRADVQVPVTFVAVVALAAMRWLGEVPAVVGSHSGEIAAHVAVLSLEDAARVYA
LRSQLARELAGRSMASVALAADVESRLAGAEAGGVRDVEIAAVNGPETTVCGA
PGAVDSLGLVLOGEGVVRIDVYASHSRHVEGI RDELAVALAGLRPRAGRVPEYST
VEAPPLDGTALDAGVYRNLRORYREFESALRAMLDGVDAFVECSPPHVLTPVROT
EDAGAGVAVGSLRDECGLRRLTSAEAQVAGVPPVDMALCPRAGWDLPTAFOR
ERYVAPAEPPGPAAGASAAATCPAATAARSGSGQEDASATGDRLATYHAWOGLP
GADAWRPGPRLLIVPAGERAEVADSVQALASFGTAVRVADPLSTGRDGLALLA
PVVAGEPLTGAVSLGLCTDGHPEHPVAGVGTALVQALADLGVAVPLMTVTRG
TVLVGSDSTVPARLLRSLDDGAEVRVLAGADAGTATDAGTAEIAPDAVVRIPC DV

Query Match 46.7%; Score 47.2; DB 1; Length 41097;

Best Local Similarity 67.0%; Pred. No. 0.034;

Matches 67; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2 TCGACGTGCGACGAGGCTTCGCGAGCA3GGCTCGACTCCCTGATGGCTGTGGAGATCC 61

Db 29776 TCGCCTCGGACCGGGCTTCGCGAGCG3GGCTTCGACTCCCTGACCGCGGTGCACTGC 29835

QY 62 GCAACGCGCTTCAGGCTGAGCTGGGTAA3CCGCTGTGCGC 101

Db 29836 GCAACGCGCTGGCGCGGCGGAGGCTTA3CGCTGTGCCGC 29875

Search completed: November 5, 2002, 13:14:17
Job time : 611.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050

Perfect score: 101
Sequence: 1 gcatcggtgcgcacacct.....tcgacgcgcagggcagcta 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.6	35.2	1600	12	BG025615
2	31.4	31.1	924	12	BG481933
3	30.8	30.5	512	10	BE481325
4	30.6	30.3	338	10	BB868343
5	30.6	30.3	921	14	BM817240
6	30.6	30.3	1101	17	CNS016HE

RESULT 1	LOCUS	DEFINITION	ACCESSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
BG025615/c	LOCUS	602274546F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4362754 5', mRNA sequence.	BG025615	EST.	human.	Homo sapiens	1 (bases 1 to 1600)	NIH-MGC	http://mhc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
BG025615.1	GI:12412391										
EST.											
human.											
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.											
1 (bases 1 to 1600)											
NIH-MGC											
http://mhc.nci.nih.gov/											
Unpublished (1999)											
Contact: Robert Strausberg, Ph.D.											
Email: cgapbs@mail.nih.gov											
Tissue Procurement: Louis Staudt, M.D., Ph.D.											
cDNA Library Preparation: Life Technologies, Inc.											
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)											
DNA Sequencing by: Incyte Genomics, Inc.											
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:											
http://image.llnl.gov											
Plate: LLM1007											
High quality sequence stop: 657.											
Location/Qualifiers											
1..1600											

FEATURES

source

[illegible]

Query Match	30.3%	Score 30.6	DB 10	Length 338
Best Local Similarity	56.4%	Pred. No. 1e+02		
Matches	57	Conservative	0	Mismatches 44; Indels 0; Gaps 0;
QY	1	GCATGGGTCCTCCACACCCCTTACCGGAGAGACCGCGCTGACCTTCCTGCTGTTTCC	60	
DB	161	GGGGGGGAGGGCGACACCATGGGCAACCGCGGGATGTAAGTAAGATCATCCCGCTGTTAAC	220	
QY	61	TCGGCGTCGGGGCGCTTCCTGCGCTCGATCGCGCAGGGACAGCA	101	
DB	221	AAGTCGACAGAGACGCGCTTCAGCTCATCGTACGTAGAGACTGCCA	261	
RESULT 5				
LOCUS	BM817240/c	921 bp	mRNA	linear
DEFINITION	HC108H04_SK.ab1	HC Hordeum vulgare	cDNA clone	HC108H04_SK.ab1, mRNA
ACCESSION	BM817240			
VERSION	BM817240.1	GI:19153254		
KEYWORDS	EST.			
SOURCE	Hordeum vulgare.			
ORGANISM	Hordeum vulgare			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae			
AUTHORS	1 (bases 1 to 921) Ozuyur,N.Z., Michalowski,C.B., Brazille,S., Borchert,C., Palacio,C., Normand,C., Murphy,C., Kelley,R., Sant,S.A., McLaughlin,H., Fredricksen,M.A. and Bohnert,H.J.			
TITLE	Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Mark A. Fredricksen Plant Biology University of Illinois 1201 W Gregory Dr, Urbana, IL 61801, USA Tel: 2172655473 Email: bohnertlab@life.uiuc.edu.			
FEATURES				
SOURCE	1..921 /organism="Hordeum vulgare" /strain="cv tokax" /db_xref="taxon:4513" /clone="HC108H04_SK.ab1" /clone_1lb="HC" /tissue_type="Root" /dev_stage="3 week old" /note="6 and 10 hour drought stress by placing plants on moist paper (75% rel. humidity) in light"			
BASE COUNT	270 a 257 c 226 g 168 t			
ORIGIN				
Query Match	30.3%	Score 30.6	DB 14	Length 921
Best Local Similarity	62.3%	Pred. No. 1.3e+02		
Matches	48	Conservative	0	Mismatches 29; Indels 0; Gaps 0;
QY	10	CTGCACACCTTACCGGAGAGCGCGCTGGACCTCTGATCGTTTCCCGGCGTCG	69	
DB	349	CTTCACACCATTTTGGCGACAGCTGCCGTTGTTGCTCGCGGAAATGTTGGCTTGCGTC	290	
QY	70	GCGCTCTTCGCTCGAT	86	
DB	289	GGTCTCGGTGGTTTGT	273	
RESULT 6				
LOCUS	CNS016HE	1101 bp	DNA	linear
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC			
ACCESSION	BA016W12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
	AT106748			

VERSION AL106748.1 GI:5623614
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-and sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : T7"
BASE COUNT 239 a 215 c 191 G 239 t 217 others
ORIGIN
Query Match 30.3%; Score 30.6; DB 17; Length 1101;
Best Local Similarity 44.7%; Pred. No. 1.4e+02;
Matches 38; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
QY 9 GCTGCACACCTTACCGCGAGCAGCCGCTCTGCTACTGTTTCTCGCGCTC 68
DB 860 GCTGCCCCCSCYSCCGSSSGSGGGGKTKTYSSTKTCCTCGCCTY 919
QY 69 GGGCGTCTTGGCTCGATCGCGCAG 93
DB 920 CBGCGGGCCYKCBCTCTGCGCTG 944
RESULT 7
BI338379/c 517 bp mRNA linear EST 30-JUL-2001
LOCUS BI338379
DEFINITION 362137 MARC 1PTG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BI338379
VERSION BI338379.1 GI:15031662
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 517)
Fahrenkrug S.C., Fekking B.A., Rohrer G.A., Smith T.P.L., Casas E.,
Stone R.T., Heaton M.P., Grosse W.M., Bennett G.A., Laegreid W.W.
and Keele J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt. trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR primers

FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCACGACG
Plate: 130 row: K column: 20
Seq primer: ATTAGGTGACTATAG.
FEATURES
source
1. 517
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1PTG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 107 a 126 c 186 g 98 t
ORIGIN
Query Match 30.1%; Score 30.4; DB 13; Length 517;
Best Local Similarity 57.3%; Pred. No. 1.3e+02;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 2 CATGGTGTGTCACACCTTACCGCGAGCAGCGCGCTGACCTGTTCTGTTTCT 61
DB 207 CTTTCTGAGGCTAGAGACCCCGAGAGGTGCTCCGACCCCTGCTCACTGTTCTCA 148
QY 62 CGCGGTGCGGCGCTTGGCTCGATCGCGCAGGCA 97
DB 147 GGGCGTGAAGGCTCTCCGCCCTTGGCTGGCGCA 112
RESULT 8
AA917322/c 379 bp mRNA linear EST 23-JUN-1998
LOCUS AA917322
DEFINITION on45b01.s1 NCI-CGAP Co8 Homo sapiens cDNA clone IMAGE:1559593 3'
similar to gb:M16538 GUANINE NUCLEOTIDE-BINDING PROTEIN G(II)/G(S
)/G(T) BETA SUBUNIT (HUMAN)), mRNA sequence.
ACCESSION AA917322
VERSION AA917322.1 GI:3057212
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 379)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cga@p5-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/bbrp/image/image.html
Trace considered overall poor quality
Insert Length: 552 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES
source
1. 379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1559593"
/clone_lib="NCI-CGAP Co8"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73b-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from

colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 100 c 110 g 71 t

Query Match 29.9%; Score 30.2; DB 9; Length 379;

Best Local Similarity 65.7%; Pred. No. 1.3e+02; Matches 44; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Oy 28 GAGCAGCCGCTCTTCTGCTACTGTTTCCGCGGCGGCGTCTGCGCTCATC 87

Db 169 GAGGAGCAGAGAGCCCTCATCTTCTGCTTCCCTGGGTTGGGCGCTCATCCCTCTGGA 110

Oy 88 GGCCAGG 94

Db 109 GGCCCGG 103

RESULT 9

LOCUS

C19416 440 bp mRNA linear EST 03-APR-2002

DEFINITION C19416 Rice panicle at ripening stage Oryza sativa (japonica cultivar-group) cDNA clone E10379, mRNA sequence.

ACCESSION C19416 C93186

VERSION C19416.2 GI:15072238

KEYWORDS

SOURCE

ORGANISM Oryza sativa (japonica cultivar-group).

Oryza sativa (japonica cultivar-group).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Embarbolidae; Oryzaceae; Oryza.

1 (bases 1 to 440)

Sasaki, T.

Rice cDNA from panicle at ripening stage

Unpublished (1996)

On Oct 24, 1996 this sequence version replaced gi:1631687

gi:13761938.

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki

305-8602, Japan

Tel: 81-298-38-7441

Fax: 81-298-38-7468

Email: tsasaki@abrr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

PROJECT = 'RGP'

RESULT 10
LOCUS B0172791 571 bp mRNA linear EST 29-APR-2002

DEFINITION B0172791 1008b05 rice seed cDNA library Oryza sativa cDNA, mRNA sequence.

VERSION B0172791.1 GI:2037161

KEYWORDS

SOURCE

ORGANISM

Oryza sativa.

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Embarbolidae; Oryzaceae; Oryza.

1 (bases 1 to 571)

Wang, X., Zhao, J., Li, Q., and Zhang, X.

Expressed sequence tags analysis of a rice seed cDNA library

Unpublished (2002)

Contact: Xiansheng Zhang

Plant development molecular biology lab

college of life sciences

Shan dong agriculture university, Tai'an, Shan dong, China

Tel: 086-0538-8241144

Email: zhiwu@sdau.edu.cn.

Location/Qualifiers

1. 571

/organism="Oryza sativa"

/db_xref="taxon:4530"

/clone_lib="rice seed cDNA library"

/tissue_type="seed"

/note="Vector: pT-Adv"

BASE COUNT 83 a 189 c 198 g 101 t

ORIGIN

Query Match 29.9%; Score 30.2; DB 14; Length 571;

Best Local Similarity 58.2%; Pred. No. 1.5e+02;

Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Oy 6 GGTGCTGCACACCTTACCGGAGACCGCTGACCTCTTCTGCTCTTCTCGGC 65

Db 293 GGTGCTGTCATCTTACCGGAGACCGCTGACCTCTTCTGCTCTTCTCGGC 352

Oy 66 GTGCGGCGCTTCTGCGCTGATCGCCAGGCG 96

Db 353 GGCCGGTTCGTACCTCGGCGGCGCTCGGC 383

RESULT 11

LOCUS BM920552/c 1146 bp mRNA linear EST 12-MAR-2002

DEFINITION BM920552 6709522 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5750558

5', mRNA sequence.

ACCESSION BM920552

VERSION BM920552.1 GI:19370931

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 1146)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bld-research.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LMU at:

http://image.llnl.gov

Plate: LLM12781 row: 1 column: 15

Query Match 29.7%; Score 30; DB 13; Length 856;
 Best Local Similarity 67.7%; Pred. No. 1.9e+02;
 Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 37 CCGACCTTCGCTGCTTTCTTCCTGCGGCTGCGGCTTCGCGCATGCGCGAGGC 96
 DB 850 CCGACCTTCGCTGCTTTCTTCCTGCTTCAGTTCAGTCTTGTTAGTGCACGCGGCG 791

QY 97 AG 98
 DB 790 AG 789

RESULT 14
 BM452552/c 988 bp mRNA linear EST 05-FEB-2002
 LOCUS BM452552
 DEFINITION AGENCOURT 6401029 NIH_MGC 67 Homo sapiens cDNA clone IMAGE:5493399
 5', mRNA sequence.
 ACCESSION BM452552
 VERSION BM452552.1 GI:18501592
 KEYWORDS EST
 SOURCE human.
 ORGANISM Homo sapiens
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 988)
 NIH-MGC http://mgc.nci.nih.gov/
 TITLE Unpublished (1999)
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs.femail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ULNL at:
 http://image.llnl.gov
 Plate: LLM12117 row: a column: 16
 High quality sequence stop: 504.
 Location/Qualifiers
 1..988
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5493399"
 /clone_lib="NIH_MGC_67"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: PCMV-SPORT6; Site: 1; Not:
 Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.75 Kb. Library constructed by Life
 Technologies."
 BASE COUNT 262 a 242 c 296 g 182 t 6 others
 ORIGIN

Query Match 29.7%; Score 30; DB 13; Length 988;
 Best Local Similarity 64.3%; Pred. No. 2e+02; Indels 0; Gaps 0;
 Matches 45; Conservative 0; Mismatches 25; Indels 0;

QY 31 CAGCCGCTGACCTTCCTGCTGCTTCGCGGCTTCGCGCTGCGCATGCGGC 90
 DB 165 CAGCTCAGAGACCGCGCTTACAGTTTACAGGCACTCGAGCGCTTTGGCGGAGCGCC 106

QY 91 CAGGCGAGCT 100
 DB 105 CGTCGCTCCT 96

RESULT 15
 AA081455/c 371 bp mRNA linear EST 23-DEC-1997
 LOCUS AA081455
 DEFINITION z004f10.s1 Stragene hmt neuron (#937233) Homo sapiens cDNA clone
 IMAGE:546475 3', mRNA sequence.

ACCESSION AA081455
 VERSION AA081455.1 GI:1623513
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 371)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
 B., Morris, M., Parsons, J., Plange, C., Rifkin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 TITLE JOURNAL
 MEDLINE
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 WARNING: There is evidence that suggests that the 384-well parent
 plate of this clone contains both human and mouse derived clones.
 Thus, the origin of this clone is uncertain. This caution should be
 kept in mind should you use this clone.

FEATURES
 source
 Location/Qualifiers
 1..371
 /organism="Homo sapiens"
 /db_xref="GDB:3924665"
 /db_xref="taxon:9606"
 /clone_image="546475"
 /clone_lib="Stragene hmt neuron (#937233)"
 /dev_stage="hmt neurons"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site: 1; EcorI; Site: 2;
 XhoI; Cloned unidirectionally. Primer: Oligo dT.
 Differentiated, post mitotic hmt neurons. Average insert
 size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
 GAATTCGGCAGAG 3', -3' adaptor sequence: 5'
 CTGAGATTGTTTTTTTTTTTTTTT 3'"
 CCGAGATTGTTTTTTTTTTTTTTT 3'"
 BASE COUNT 125 a 68 c 81 g 97 t
 ORIGIN

Query Match 29.5%; Score 29.8; DB 9; Length 371;
 Best Local Similarity 70.2%; Pred. No. 1.7e+02;
 Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 CATGGTGTGACACCTTACCGCGAGCAGCGCTGACCTCTGCTGCTTTT 58
 DB 360 CAAGTTGCTGACCTCATACAGCTGCTCCTCTGCTGCTCTGCTGCTTTT 304

Search completed: November 5, 2002, 16:17:40
 Job time: 119.44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050

Perfect score: 101

Sequence: 1 gcattggtcgtcgcacacccct.....tcgattcgccagggcagcta 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_NA:
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	68750	3	US-09-335-409-1
2	101	100.0	68750	4	US-09-568-102-1
3	101	100.0	68750	4	US-09-567-969-1
4	101	100.0	68750	4	US-09-568-480-1
5	101	100.0	68750	4	US-09-568-486-1
6	101	100.0	68750	4	US-09-568-472-1
7	101	100.0	68750	4	US-09-567-899-1
8	101	100.0	71989	4	US-09-443-501A-2
9	49.8	49.3	15872	4	US-09-105-537-1
10	41.8	41.4	4403765	4	US-09-103-840A-2
11	41.8	41.4	4411529	4	US-09-103-840A-1
12	40.8	40.4	33529	4	US-09-144-085-3
13	40.8	40.4	80161	4	US-09-036-987A-1
14	40.8	40.4	80161	4	US-09-370-700-1
15	40.2	39.8	4689	4	US-09-105-537-34
16	40.2	39.8	13842	4	US-09-105-537-30
17	40.2	39.8	36778	4	US-09-105-537-5
18	40.2	39.8	38506	3	US-09-320-878-19
19	37.6	37.2	11219	1	US-07-642-734C-1
20	37.6	37.2	11219	3	US-08-439-009A-1
21	36.2	35.8	13987	2	US-08-804-227C-13
22	36.2	35.8	44377	2	US-08-804-227C-7
23	36.2	35.8	44377	2	US-08-804-198-1
24	35.4	35.0	28958	1	US-08-258-261B-6
25	35.4	35.0	28958	1	US-08-456-837-6
26	35.4	35.0	28958	1	US-08-457-342-6
27	35.4	35.0	28958	1	US-08-457-646A-6

28	35.4	35.0	28958	1	US-08-458-076A-6	Sequence 6, Appli
29	35.4	35.0	28958	1	US-08-764-233A-4	Sequence 4, Appli
30	35.4	35.0	28958	1	US-08-457-335A-6	Sequence 6, Appli
31	35.4	35.0	28958	1	US-08-729-214-6	Sequence 6, Appli
32	35.4	35.0	28958	3	US-09-028-934-6	Sequence 6, Appli
33	35.4	35.0	49377	1	US-08-764-233A-1	Sequence 1, Appli
34	34.4	34.1	43280	2	US-08-804-227C-1	Sequence 1, Appli
35	34	33.7	1681	4	US-09-434-288-7	Sequence 7, Appli
36	33	32.7	50937	4	US-09-428-517-1	Sequence 1, Appli
37	31.4	31.1	11220	4	US-09-105-537-32	Sequence 32, Appli
38	31.2	30.9	477	3	US-09-010-809-2	Sequence 2, Appli
39	31.2	30.9	477	3	US-09-010-809-17	Sequence 17, Appli
40	31	30.7	1601	1	US-08-206-006-1	Sequence 1, Appli
41	30.6	30.3	20235	1	US-07-642-734C-3	Sequence 3, Appli
42	30.6	30.3	20235	3	US-08-439-009A-3	Sequence 3, Appli
43	30.4	30.1	11561	1	US-08-450-332-1	Sequence 1, Appli
44	30.4	30.1	11561	2	US-08-637-640-1	Sequence 1, Appli
45	30.4	30.1	11561	4	US-09-004-406C-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1

Query Match      100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCATGGGCTCTGCACACCCCTTACCGCGAGCAGCCGCTTGACTTCTTTC 60
        |||||||
DB 2562 GCATGGGCTCTGCACACCCCTTACCGCGAGCAGCCGCTTGACTTCTTTC 25621

QY      61 TCGGGGTGCGGGGCTTGGGCTCGATGGCCGAGGGCAGCTA 101
        |||||||
DB 2562 TCGGGGTGCGGGGCTTGGGCTCGATGGCCGAGGGCAGCTA 25622

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
```

```
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 60
Db 25562 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 25621
```

```
OY 61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 101
Db 25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 25662
```

```
RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 60
Db 25562 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 25621
```

```
OY 61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 101
Db 25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 25662
```

```
RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
```

```
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1
```

```
Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 60
Db 25562 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 25621
```

```
OY 61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 101
Db 25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 25662
```

```
RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1
```

```
Query Match 100.0%; Score 101; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 60
Db 25562 GCATGGGTGCTGCACACCCCTTACCCGCGAGCAGCGGTGACCTTTCGTACTGTTTCC 25621
```

```
OY 61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 101
Db 25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCGCAGGACAGCTA 25662
```

```
RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
```


FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 40.4%; Score 40.8; DB 3; Length 80161;
Best Local Similarity 65.2%; Pred. No. 0.0014;
Matches 60; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 10 CTGCACACCTTACCCGCGAGCAGCCGCTGACCTCTTCGTACTGTTTCTCGGCCTCG 69
DB 58799 CTGACCGCGCTGCTTGGCGGCGACACCGCTTGACGCTTGTCTCTCATCTCG 58858

QY 70 GGCGTCTTGGCTCGATCGGCGCAGGCGACTA 101

DB 58859 GGAGTTTGGGAGCGCGCGCCAGAGGCTTA 58890

RESULT 14

US-09-370-700-1

; Sequence 1, Application US/09370700

; Patent No. 6274350

; GENERAL INFORMATION:

; APPLICANT: Baltz, Richard H

; APPLICANT: Broughton, Mary C

; APPLICANT: Crawford, Kathryn P

; APPLICANT: Madduri, Krishnamurthy

; APPLICANT: Treadway, Patli J

; APPLICANT: Turner, Jan R

; APPLICANT: Waldron, Clive

; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide

; FILE REFERENCE: 50489 Div1

; CURRENT APPLICATION NUMBER: US/09/370,700

; CURRENT FILING DATE: 1999-08-09

; EARLIER APPLICATION NUMBER: US 09/36987

; EARLIER FILING DATE: 1998-03-09

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 80161

; TYPE: DNA

; ORGANISM: Saccharopolyspora spinosa

US-09-370-700-1

Query Match 40.4%; Score 40.8; DB 4; Length 80161;
Best Local Similarity 65.2%; Pred. No. 0.0014;
Matches 60; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 10 CTGCACACCTTACCCGCGAGCAGCCGCTGACCTTCTGTAAGTTTCTCGGCCTCG 69

DB 58799 CTGACCGCGCTGCTTGGCGGCGACACCGCTTGACGCTTGTCTCTCATCTCG 58858

QY 70 GGCGTCTTGGCTCGATCGGCGCAGGCGACTA 101

DB 58859 GGAGTTTGGGAGCGCGCGCCAGAGGCTTA 58890

RESULT 15

US-09-105-537-34

; Sequence 34, Application US/09105537A

; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-34

Query Match 39.8%; Score 40.2; DB 4; Length 4689;
Best Local Similarity 62.4%; Pred. No. 0.0014;
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGCAGCCGCTGACCTCTTCGTACTGTTTTC 60
DB 3688 GCCGAGTCTCTGACGACCTGCTCGCGGACACTCCGTGACGCTTGTCTCTACTCC 3747

QY 61 TCGGCGTGGGCGCTTTCGCTCGATCGGCGCAGGCGACTA 101

DB 3748 TCGAAGCGCGGGGCTGTGGGGCAGCGGCGACGAGGCGCTTA 3788

Search completed: November 5, 2002, 14:05:11
Job time : 1655.74 secs


```
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-34
```

```
Query Match          39.8%; Score 40.2; DB 10; Length 4689;
Best Local Similarity 62.4%; Pred. No. 0.00078;
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
Oy 1 GCATGGGTGCTGCACACCTTACCGGAGACAGCGGCTGGACCTCTCGTACTGTTTCC 60
    || ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3688 GCCAGGTCTCTCGACGACGCTGCTCGCGGAACTCGCTGGACGCTTCTCTACTCC 3747

Oy 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCACGGCAGCTA 101
    || ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3748 TCGAACGCCGGGGCTCTGGGGCAGCGCAGCAGCGGCGTCTA 3788
```

```
RESULT 3
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30
```

```
Query Match          39.8%; Score 40.2; DB 10; Length 13842;
Best Local Similarity 62.4%; Pred. No. 0.00088;
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
Oy 1 GCATGGGTGCTGCACACCTTACCGGAGACAGCGGCTGGACCTCTCGTACTGTTTCC 60
    || ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6864 GCCTCGGTCTCGACGACGCTGACCCGGGACCTCGACCTCGACGCGTCTCTCTCG 7023

Oy 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCACGGCAGCTA 101
    || ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7024 TCCGTGTGAGCACTCTGGGCATCCCGGACAGGCAACTA 7064
```

```
RESULT 4
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
```

```
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match          39.8%; Score 40.2; DB 10; Length 36778;
Best Local Similarity 62.4%; Pred. No. 0.00099;
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
Oy 1 GCATGGGTGCTGCACACCTTACCGGAGACAGCGGCTGGACCTCTCGTACTGTTTCC 60
    || ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8705 GCCTCGTCTCGACGACGCTGACCCGGGACTCGACTCGACGCGTCTCTCTCG 8764

Oy 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCACGGCAGCTA 101
    || ||||| || ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8765 TCCGTGTGAGCACTCTGGGCATCCCGGACAGGCAACTA 8805
```

```
RESULT 5
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32
```

```
Query Match          31.1%; Score 31.4; DB 10; Length 11220;
Best Local Similarity 59.6%; Pred. No. 0.42; 36; Indels 0; Gaps 0;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
Oy 13 CACACCTTACCGGAGACAGCGGCTGACCTCTTCGTAATGTTTCCTCGCGCGGCG 72
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10192 CTGACCTCGACGCGCGGCTCGACCTGCGACGCTTCTGTAATGTTTCCTCGCGCGGCG 10251

Oy 73 GCTCTGCTCGATCGGCACGGCAGCTA 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10252 GTCTTGGTGGCGCGGCGGACAGGCGCCTA 10280
```

```
RESULT 6
US-09-815-242-7877
; Sequence 7877, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
```

```

; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7877
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1059)
; US-09-815-242-7877

Query Match          28.3%; Score 28.6; DB 10; Length 1059;
Best Local Similarity 57.1%; Pred. No. 2.3;
Matches 52; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```

QY 6 GGTGTCGACACCTTACCCGCGAGCAGCCGCTGACCTCTTCGACTGTTTCCCTCGGC 65
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 792 GGTGTCGAGGAGCTGAGGAGGTGAGGTGAGGCTGACGCTTCGACTGTCACCGCA 851
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 GTCGGCGCTCTCGGCTGATCGGCGCAGGGC 96
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 852 ACTGCGCTCTCGGCGAGCGCTGACCGC 882

RESULT 7
US-09-294-093B-3349
; Sequence 3349, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalugudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 3349
; LENGTH: 299
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700379878H1
; NAME/KEY: unsure
; LOCATION: 60, 214, 278
```

```

; OTHER INFORMATION: a, t, c, g, or other
; US-09-294-093B-3349
```

```

Query Match          28.1%; Score 28.4; DB 10; Length 299;
Best Local Similarity 59.5%; Pred. No. 2.3;
Matches 47; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```

QY 18 CCTTACCGCGAGCAGCCGCTGACCTCTTCGACTGTTTCTCGGCGTCGGCTCTT 77
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 215 CTTACGCGCGAGCGGCGCCCGCCAGTTCTCGGTGGCGGTCTCTCTCAGGCTCAT 274
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 78 CGGCTCGATCGGCGCAGGGC 96
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 275 CGANCCGGCAGCCCTGGAC 293
```

```

RESULT 8
US-09-777-564-12/c
; Sequence 12, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 12
; LENGTH: 219
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-777-564-12
```

```

Query Match          27.7%; Score 28; DB 10; Length 219;
Best Local Similarity 60.5%; Pred. No. 2.9;
Matches 46; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```

QY 15 CACCTTACCGCGAGCAGCCGCTGACCTCTTCGACTGTTTCTCGGCGTCGGGCT 74
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 88 CACACTGACCGCGGCTCCACCTTGCGCCCTCGAGCGTACTGATTTCTCTCATGTTCT 29
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 CTTGCGCTGATCGGC 90
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 28 CTTGAGGTAGGCCAGC 13
```

```

RESULT 9
US-09-777-564-1473/c
; Sequence 1473, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Mannion, Jane
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 1473
; LENGTH: 219
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-777-564-1473
```

```

Query Match          27.7%; Score 28; DB 10; Length 219;
Best Local Similarity 60.5%; Pred. No. 2.9;
Matches 46; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

[illegible]

```

RESULT 10
US-09-764-887-133
: Sequence 133. Application US/09764887
: Patent No. US200200642096A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PAl13
: CURRENT APPLICATION NUMBER: US/09/764.887
: PRIOR APPLICATION FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 658
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 133
: LENGTH: 518
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-887-133

```

Query Match	27.3%	Score 27.6	DB 10	Length 518
Best Local Similarity	63.6%	Pred. No. 4.3		
Matches 42; Conservative	0;	Mismatches 24;	Indels 0;	Gaps 0;

Oy 35 CGCAGGACCTTCGTACTGTTTCCCTGG;CGTGGGCGCTCTCGGCTCAATGGGCCAGG 94
 ||| |||| | ||| |||| | ||| ||| |||| ||| |||
Db 258 CGCAGGACCGCGCTTTACAGTTTACCAGG;ACTCGGACGCCTTTGGCGCGAGCGCCGTC 317

Qy	95	GCAGCT	100
Db	318	GCTCCT	323

```

RESULT 11
US-09-764-887-603
: Sequence 603: Application US/09764887
: Patent No. US20020042096A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P413
: CURRENT APPLICATION NUMBER: US/09/764,887
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 658
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 603
: LENGTH: 518
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-764-887-603

```

Query Match	27.3%	Score 27.6	DB 10	Length 518
Best Local Similarity	63.68	Pred. No. 4.3		
Matches 42	Conservative 0	Mismatches 24	Indels 0	Gaps 0

Dy 35 CGCTGGACCTTCGTAAGTATTTTCCTCGCCGTTCGGCTCATCGGCCAGG 94
||| |||| | ||| |||| | ||| |||| | ||| ||||
Db 258 CCGAGGACCGCGCTTACAGTTACCAAGCACTCGGACGCTTGCGCGCGAGCGCCGTC 317

QY	95	GCAGCT	100
Db	318	GCTCCT	323

```

RESULT 12
US-09-764-887-604
: Sequence 604, Application US/09764887
: Patent No. US20020042096A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P113
: CURRENT APPLICATION NUMBER: US/09/764, 887
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 658
: SOFTWARE: PatentVer. 2.0
: SEQ ID NO 604
: LENGTH: 518
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-887-604

```

Query Match	27.3%	Score 27.6	DB 10	Length 518
Best Local Similarity	63.6%	Pred. No. 4.3		
Matches 42; Conservative	0	Mismatches 24	Indels 0	Gaps 0

Qy 35 CGCTGGACCTTTCGACTGTATCCCGGGCGGCTTGAGTCGGGCAGG 94
||| ||||| - ||| ||||| - |||| | ||| ||| ||| ||
Db 258 CCGCAGCACCGCGCTTTACAGTTTACCAGGCACTCGGACGCCCTTTGGCGGCGAGCGCCCGTC 317

QY	95	GCAGCT	100
Db	318	GCTCCT	323

```

RESULT 13
US-09-764-887-605
: Sequence 605. Application US/09764887
: Patent No. US20020042096A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P413
: CURRENT APPLICATION NUMBER: US/09/764.887
: PRIORITY FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 658
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 605
: LENGTH: 519
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-764-887-605

```

Query Match	27.38%	Score 27.6	DB 10	Length 519
Best Local Similarity	63.68%	Pred. No. 4.3		
Matches 42; Conservative	0;	Mismatches 24;	Indels 0;	Gaps 0;

Qy 35 CGGAGGACCTTCGTACTGTTTCCCGGGCGCTCTTCGGCTCGATCGGCCAGG 94
||| ||||| | ||||| ||| ||| ||| |||
Db 259 CGCAGGACC GCGCTTTACAGTTTACCAAGGCACCTCGAGCCCTTGSGGCGCAGCGCCCGTC 318

QY	95	GCAGCT	100
Db	319	GCTCCT	324

US-RESULT 14 US/09983965
: 09-983-965-5085/c
: Sequence 5085, :
: Patient No. US20020137160A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Ito, Nengding
: APPLICANT: Byatt, John C.
: APPLICANT: Mathalagan, Nagappan

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050

Perfect score: 101
Sequence: 1 gcattgctgcgcacacct.....tcgacgcgcagcagcagcta 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N.Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	68750	21	AAZ55887
2	101	100.0	71989	21	AAAZ29349
3	49.8	49.3	15872	21	AAZ87283
4	48.2	47.7	125401	22	AAAD17186
5	46.6	46.1	552	24	AAAF68063
6	46.6	46.1	6462	22	AAAF90037
7	46.6	46.1	15872	18	AAAF68715
8	46.6	46.1	34071	22	AAAF90033
9	46.6	46.1	42717	22	AAAF90032

10	45.6	45.1	552	24	AAI68064	9-keto-epothillone
11	45.6	45.1	5676	19	AAV21186	Amycolatopsis medi
12	45.6	45.1	27541	22	AAAD17185	Streptomyces nous
13	45.6	45.1	53789	19	AAV21187	Amycolatopsis medi
14	44	43.6	6333	22	AAAF31641	Mycobacterium tube
15	44	43.6	8301	22	AAAF90035	Nucleotide sequenc
16	43.4	43.0	20394	22	AAAF24892	Pimaric acid biosynth
17	43.4	43.0	30690	21	AAAF92301	S. avermilitis ave
18	43.4	43.0	30690	22	AAAF92301	Streptomyces averm
19	43.4	43.0	65140	22	AAAD17184	Streptomyces nous
20	43.4	43.0	77536	21	AAAF14651	Nucleotide sequenc
21	41.8	41.4	47981	22	AAAF30757	Micromonospora meg
22	41.8	41.4	4403765	22	AAAF9683	Mycobacterium tube
23	41.8	41.4	4411529	22	AAAF9682	Streptomyces averm
24	41.2	40.8	11916	22	AAAF79279	Streptomyces averm
25	41.2	40.8	12381	21	AAZ58381	Streptomyces averm
26	41.2	40.8	31422	21	AAAF92302	S. avermilitis ave
27	41.2	40.8	31422	22	AAAF92302	Streptomyces averm
28	40.8	40.4	14775	22	AAAF88338	S. spinosa DNA fra
29	40.8	40.4	33529	23	AAAF17367	DNA sequence of S.
30	40.8	40.4	50000	22	AAAF88313	S. spinosa DNA fra
31	40.8	40.4	50000	22	AAAF88316	S. spinosa DNA fra
32	40.8	40.4	80161	20	AAZ21501	DNA fragment of Sa
33	40.2	39.8	4466	21	AAAF14663	Nucleotide sequenc
34	40.2	39.8	4478	21	AAAF14661	Nucleotide sequenc
35	40.2	39.8	4547	21	AAAF14664	Nucleotide sequenc
36	40.2	39.8	4571	21	AAAF14662	Nucleotide sequenc
37	40.2	39.8	4689	21	AAZ87299	S. venezuelae macr
38	40.2	39.8	5088	22	AAAF90038	Nucleotide sequenc
39	40.2	39.8	5292	22	AAAF90036	S. venezuelae macr
40	40.2	39.8	13842	21	AAZ87297	S. venezuelae pik
41	40.2	39.8	36778	21	AAZ87318	S. venezuelae pik
42	40.2	39.8	37948	21	AAZ87285	AAAF25633
43	40.2	39.8	38506	21	AAZ56001	AAAF25603
44	40.2	39.8	38506	21	AAZ56001	AAAF25603
45	40.2	39.8	77536	21	AAAF14651	Nucleotide sequenc

ALIGNMENTS

RESULT 1	AAZ55887	standard; DNA; 68750 BP.
ID	AAZ55887	
AC	AAZ55887	
XX		
DT	10-APR-2000	(first entry)
XX		
DE	Sorangium cellulosum 68.75 kb contig.	
XX		
KW	Epithilone biosynthesis; type I polyketide synthase; taxol substitute;	
KW	anticancer; ds.	
XX		
OS	Sorangium cellulosum.	
XX		
FH	key	Location/Qualifiers
FT	CDS	1..1626
FT		/*tag= a
FT		/partial
FT		/product= "Partial Orf 1 protein (AAV58580)"
FT		/note= "No initiation codon given in the specification"
FT	CDS	complement (1900..3171)
FT		/*tag= b
FT		/product= "Orf 2 protein (AAV58581)"
FT	CDS	3415..5556
FT		/*tag= c
FT		/product= "Orf 3 protein (AAV58582)"
FT	CDS	complement (5612..5992)
FT		/*tag= d
FT		/product= "Orf 4 protein (AAV58583)"
FT	CDS	6226..6675
FT		/*tag= e

```

FT      /product= "Orf 5 protein (AAV58584)"
FT      7610..11875
FT      /*tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
FT      /*tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /*tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /*tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /*tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /*tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /*tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /*tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /*tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /*tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /*tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /*tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /*tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /*tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /*tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /*tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /partial
FT      /*tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX      MO9966028-A2.
XX      PD
XX      23-DEC-1999.
XX      PF
XX      16-JUN-1999; 99WO-EP04171.
XX      PR
XX      18-JUN-1998; 98US-009504.
XX      24-SEP-1998; 98US-0101631.
XX      05-FEB-1999; 99US-0118906.
XX      PA (NOVS ) NOVARTIS AG.

```

```

PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERM GBS MBH.
XX      PI
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX      WPI: 2000-097741/08.
DR      P-PSDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
DR      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
DR      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
DR      AAV58592, AAV58593, AAV58594.
XX      PT
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy.
XX      PS
XX      Claim 14; Page 87-104; 174pp; English.
CC      This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC      comprising 22 open reading frames (ORFs) and includes genes encoding
CC      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC      starter unit; polyketides being synthesised from two-carbon building
CC      blocks, the beta-carbon of which always carries a keto group. Each round
CC      of two-carbon addition is carried out by a complex of enzymes known as
CC      the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
CC      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
CC      and EPOS E (AAV58575-Y58578) are involved in polyketide backbone
CC      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
CC      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
CC      involved in transport. Epoethiones mimic the biological activity of
CC      taxol, and may be substituted for taxol in cancer chemotherapeutic
CC      compositions. Epoethiones exhibit a much lower drop in potency against a
CC      multiply drug-resistant cell line compared with taxol, and are
CC      considerably less efficiently exported from such cells by the multidrug
CC      resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC      epoethiones as anticancer agents, they are problematical to produce on a
CC      large scale. Epoethiones are too complex for industrial scale chemical
CC      synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC      poor yields of epoethiones. The nucleic acids of the invention may be
CC      used for the recombinant production of epoethiones in a heterologous host
CC      that is more amenable to fermentation.
XX      SQ
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other:
XX      Query Match 100.0%; Score 101; DB 21; Length 68750;
XX      Best Local Similarity 100.0%; Pred. No. 6..5e-21;
XX      Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GCATGGGTGCTGCACACCTTACCCGCGAGACGCGCTGACCTCTTGTACTTTTCC 60
DB      25562 GCATGGGTGCTGCACACCTTACCCGCGAGACGCGCTGACCTCTTGTACTTTTCC 25621
QY      61 TCGGCGTGGGCGCTTCGCGCTGCATGCGCCAGGACGACTA 101
DB      25622 TCGGCGTGGGCGCTTCGCGCTGCATGCGCCAGGACGACTA 25662
XX      RESULT 2
XX      AAA29349
XX      ID AAA29349 standard; DNA: 71989 BP.
XX      AC AAA29349;
XX      DT
XX      12-SEP-2000 (first entry)
XX      DE Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX      KW Epoethione; polyketide synthase; epOA; epOB; epOC; epOD; epOE; epOF;
XX      epOL; epOK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
XX      tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX      OS Sorangium cellulosum.
XX      FH
XX      Key Location/Qualifiers

```



```
FT CDS 3..992
FT /tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT 989..1501
FT /tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT 1998..6263
FT CDS
FT /label= epca_gene
FT /note= "encodes the loading domain"
FT 2031..3548
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT 4917..5810
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT 5856..6155
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT 6260..10493
FT /tag= h
FT /label= epob_gene
FT /note= "encodes module 1, the NRPS module"
FT 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature sequence"
FT 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT 9090..9179
FT misc_RNA

FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT 10639..16137
FT CDS
FT /tag= x
FT /label= epoc_gene
FT /note= "encodes module 2"
FT 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT 14962..15736
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT 16134..37907
FT /tag= ad
FT /label= epod_gene
FT /note= "encodes modules 3-6"
FT 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT 27966..28574
FT /tag= ao
FT /note= "encodes DH5"
FT 29433..30287
FT /tag= ap
FT misc_RNA
```

```
FT      /note= "encodes EF5"
FT      30321..30869
FT      /tag= aq
FT      /note= "encodes KF5"
FT      31077..31373
FT      /tag= ar
FT      /note= "encodes ACP5"
FT      31440..32807
FT      /tag= as
FT      /note= "encodes KE6"
FT      33018..34067
FT      /tag= at
FT      /note= "encodes AT6"
FT      34107..34676
FT      /tag= au
FT      /note= "encodes DH6"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes ER6"
FT      36705..37256
FT      /tag= av
FT      /note= "encodes KR6"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= "epoE_gene"
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT7"
FT      41341..41922
FT      /tag= db
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 6.5e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCATGGGTGCTGCACACCCCTTACCCGGCA;CAGCCGCTGACACTCTTCTGACTGTTTCC 60
DB      19950 GCATGGGTGCTGCACACCCCTTACCCGGCA;CAGCCGCTGACACTCTTCTGACTGTTTCC 20009
QY      61 TCGGCGTCGGGCGCTTCTGCGCTGCATCGG;CAGGGCAGCTTA 101
DB      20010 TCGGCGTCGGGCGCTTCTGCGCTGCATCGG;CAGGGCAGCTTA 20050

RESULT 3
AA287283
ID      AA287283 standard; DNA; 15872 BP.
XX      AA287283;
AC      AA287283;
XX      05-JUN-2000 (first entry)
DE      S. venezuelae vep ORF 1, SEQ ID NO 1.
XX      Desosamine biosynthesis: macrolide; polyketide; methymycin; pikromycin;
KM      neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KM      biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KM      chronic obstructive pulmonary disease; respiratory inflammation;
KW      hypercholesterolemia; crop protection agent; ds.
OS      Streptomyces venezuelae ATCC15439.
XX      Streptomyces venezuelae ATCC15439.
FH      Key      Location/Qualifiers:
FT      CDS      20..13912
```

```
FT      /tag= a
FT      /product= "vep ORF 1 amino acid sequence #1 (AA77177)"
FT      14056..14151
FT      /tag= b
FT      /product= "vep ORF 1 amino acid sequence #3 (AA77199)"
FT      14167..15827
FT      /tag= c
FT      /product= "vep ORF 1 amino acid sequence #2 (AA77178)"
FT      WO200000620-A2.
XX      06-JAN-2000.
XX      25-JUN-1999; 99WO-US14398.
XX      26-JUN-1998; 98US-0105537.
XX      (MINU ) UNIV MINNESOTA.
XX      Sherman DH, Liu H, Xue Y, Zhao L;
XX      WPI: 2000-160679/14.
XX      P-PSDB: AA77177, AA77178, AA77199.
XX      Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX      synthesis of methymycin and pikromycin -
XX      Example 3; Figure 23; 438pp; English.
XX      The invention relates to an isolated and purified nucleic acid segment
XX      comprising a desosamine biosynthetic gene cluster, a fragment or its
XX      biologically active variant, where the nucleic acid sequence is not
XX      derived from the ercC gene cluster of Saccharopolyspora erythraea or
XX      Streptomyces antibiotics. The invention also relates to a macrolide
XX      biosynthetic gene cluster, or fragments thereof. The macrolide
XX      biosynthetic gene cluster encodes proteins which synthesise methymycin,
XX      pikromycin, neomethymycin, narbomycin or a combination of these
XX      compounds. Recombinant or augmented cells comprising the desosamine
XX      and/or macrolide biosynthetic gene clusters are useful for the
XX      production of biologically active macrolides. The macrolide biosynthetic
XX      proteins are useful for synthesis of methymycin, pikromycin,
XX      neomethymycin and narbomycin. The alternative termination of polyketide
XX      synthesis may be useful to prepare novel antibiotics and
XX      polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
XX      recombinant host cells are useful as biopolymers, e.g., in packaging or
XX      biomedical applications, to engineer PHA monomer synthases or to prepare
XX      biologically active agents, such as chemotherapeutics,
XX      immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
XX      disease as well as other diseases involving respiratory inflammation,
XX      cholesterol-lowering agents or macrolide-based antibiotics which are
XX      active against a variety of organisms, e.g., bacteria, including
XX      multi-drug resistant pneumococci and other respiratory pathogens, as well
XX      as viral parasitic pathogens, or as crop protection agents (e.g.,
XX      fungicides or insecticides) via expression of polyketides in plants. The
XX      present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
XX      sequence, designated vep ORF 1 in the specification, which actually
XX      contains 3 open reading frames, which encode proteins AA77177-77178 and
XX      AA77199. The vep ORF 1 protein is defined in the specification as a PHA
XX      monomer synthase.
XX      Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;
SQ      5Q

Query Match      49.3%; Score 49.8; DB 21; Length 15872;
Best Local Similarity 68.3%; Pred. No. 1.4e-05;
Matches 69; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      1 GCATGGGTGCTGCACACCCCTTACCCGGAGACCGCTGACACTCTTCTGACTGTTTCC 60
DB      7682 GCGTGAACCTGCACGAGCTCACCAAGACCTCGGCGCTCGACTCTTCTCTCC 7741
QY      61 TCGGCGTCGGGCGCTTCTGCGCTGCATCGG;CAGGGCAGCTTA 101
DB      7742 TCGGTCGCGCATCGTCGCGCACCGCGCCAGGCCAAGCTTA 7782
```

```

RESULT 4
AADI17186
ID AADI17186 standard; DNA: 125401 BP.
XX
XX AADI17186;
AC
XX
XX 29-NOV-2001 (first entry)
DT
XX
DE Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KM antifungal; antibiotic; ds.
XX
OS Streptomyces noursei.
XX
XX Key Location/Qualifiers
FH 6337..34771
FT /*tag= a
FT /product= "NysI complete protein"
FT 34792..51099
FT /*tag= b
FT /product= "NysJ protein"
FT 51155..57355
FT /*tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /*tag= d
FT /product= "NysL protein"
FT complement (58786..58980)
FT /*tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement (59045..60241)
FT /*tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement (60238..61296)
FT /*tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /*tag= h
FT /product= "NysR4 (long) protein"
XX
XX WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNFE) SINTEF STITTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZORC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (ETAE/) FJAEVRIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI: 2001-557614/62.
XX P-PSDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX AAE10149, AAE10150.
XX

```

```

PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
XX Claim 1; Page 188-254; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
SQ
Query Match 47.7%; Score 48.2; DB 22; Length 125401;
Best Local Similarity 67.3%; Pred. No. 5.5e-05;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGCAGCCGCTGACCTTCTGACTGTTTCC 60
DB 55745 GCCTGGCATCTGCACGAGCCAGCCGCGGCTGACCTGCGGCTTGCCTGTACTGCC 55804
QY 61 TCGGCGCTCGGCGCTCTTCGCTCGATCGGCGCAGGCGACTA 101
DB 55805 TCGGTCTCCGCGCTCTCTGCGCAGCCCGCGCAGGCGCACTA 55845
RESULT 5
AAI68063
ID AAI68063 standard; DNA: 552 BP.
XX
XX AAI68063;
AC
XX
XX 13-MAR-2002 (first entry)
DT
XX
XX Epothilone PKS KR domain nucleotide sequence.
XX
XX Cystobacterineae; recombinant; polyketide synthase; PKS; polyketide;
XX epothilone; pila gene; cytostatic; antiproliferative; antitumor;
XX antitartaric; antitumor; antitumor; antitumor; antitumor; antitumor;
XX Synthetic.
XX
XX Key Location/Qualifiers
FH 1..552
FT /*tag= a
FT /note= "KR domain of extender module 6"
XX
XX WO200183800-A2.
XX
XX 08-NOV-2001.
XX
XX 26-APR-2001; 2001WO-US13793.
XX
XX 28-APR-2000; 2000US-0560367.
XX 14-SEP-2000; 2000US-232696P.
XX 21-DEC-2000; 2000US-257517P.
XX 03-APR-2001; 2001US-0825856.
XX 03-APR-2001; 2001US-0825876.
XX 13-APR-2001; 2001US-269020P.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Arslanian RL, Ashley G, Fryman S, Julien B, Katz L, Khosla C;
XX Lau J, Licardi PJ, Regentin R, Santi D, Tang L;
XX
XX WPI: 2002-075167/10.
XX P-PSDB: ABB07167.
XX
XX Recombinant host cells useful for producing polyketides e.g. epothilone
XX or its derivatives, comprises a recombinant expression vector encoding
XX a heterologous polyketide synthase gene -
XX
XX Example 11; Page 164-165; 221pp; English.
XX

```


DR P-PSDB; AAM19629-30 AND AAM00918.
XX Expression cassettes for production of polyhydroxyalkanoate(s) -
PT provide wide range of biodegradable polymers for medical or
PI industrial use
XX
XX Claim 54; Fig 23; 91pp; English.
XX
CC Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC synthase (PKS) gene cluster encoding a polypeptide of 12 carbons (see
CC also AAM19629-30 and AAM00918). It contains 5 PKS modules, with a 5'
CC loading module and a 3' end domain. Each of the sequenced modules
CC includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC reductase and an acyl carrier protein domain. The gene cluster was
CC cloned using a heterologous hybridisation strategy from a genomic
CC DNA library. A novel expression cassette encoding the first module
CC from the vep gene cluster and module 7 from the Streptomyces ty1p
CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC activity and can be used for PHA prodn. in host (esp. insect) cells
CC for use as a biodegradable polymer.
XX
SQ Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;
XX
Query Match 46.1%; Score 46.6; DB 18; Length 15872;
Best Local Similarity 66.3%; Pred. No. 0.00013;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
XX
OY 1 GCATGGGCTGTCACACCTTACCCGCGAGACGCCCTGACCTCTTCTGTTTTC 60
DB 7682 GCGTGAACCTGCACAGCTCACCAAGACCTGCGCTGACGCTTCTCGCTCTCC 7741
|| ||| ||||| || ||| || ||| ||| ||||| || |||
OY 61 TCGGCGTGGGCGCTCTTGGCTCGATCGGCGAGGCGAGCTA 101
DB 7742 TCCGCTCCGCGCATCTCGGCGACCGCGCCAGGCCACTA 7782
|| ||| || ||| || ||||| || ||||| || |||
XX
RESULT 8
AAF90033
ID .AAF90033 standard; DNA; 34071 BP.
XX
XX AAF90033;
XX
XX 06-AUG-2001 (first entry)
XX
XX Nucleotide sequence of cosmid a26g1 (coding strand).
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX Synthetic.
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-FR03311.
XX
XX 29-NOV-1999; 99FR-0015032.
XX
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVET) AVENTIS PHARMA SA.
XX
XX Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
XX Capellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphille K;
XX Prostegard A;
XX
XX WPI; 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics -
XX
XX Example 14; Page 289-300; 356pp; French.
XX

CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthases) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand, which encodes different
XX polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
XX
Query Match 46.1%; Score 46.6; DB 22; Length 34071;
Best Local Similarity 66.3%; Pred. No. 0.00014;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
XX
OY 1 GCATGGGCTGTCACACCTTACCCGCGAGACGCCCTGACCTCTTCTGTTTTC 60
DB 23846 GCGTGAACCTGCACAGCTCACCAAGACCTGCGCTGACGCTTCTCGCTCTCC 23905
|| ||| ||||| || ||| || ||| ||| ||||| || |||
OY 61 TCGGCGTGGGCGCTCTTGGCTCGATCGGCGAGGCGAGCTA 101
DB 23906 TCCGCTCCGCGCATCTCGGCGACCGCGCCAGGCCACTA 23946
|| ||| || ||| || ||||| || ||||| || |||
XX
RESULT 9
AAF90032/c
ID .AAF90032 standard; DNA; 42717 BP.
XX
XX AAF90032;
XX
XX 06-AUG-2001 (first entry)
XX
XX Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
XX Synthetic.
XX
XX WO200140497-A2.
XX
XX 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-FR03311.
XX
XX 29-NOV-1999; 99FR-0015032.
XX
XX 07-JUN-2000; 2000US-0209800.
XX
XX (AVET) AVENTIS PHARMA SA.
XX
XX Jeanin P, Pernodet J, Guerin M, Simonet P, Courtois S;
XX Capellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphille K;
XX Prostegard A;
XX
XX WPI; 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics -
XX
XX Example 14; Page 274-288; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX passing nucleic acid-enriched fractions through an anion exchange
XX chromatography material; and recovering fractions containing purified
XX

CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC encodes different polyketide synthases.

XX Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;

Query Match 46.1%; Score 46.6; DB 22; Length 42717;
Best Local Similarity 66.3%; Pred. No. 0.00014;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGCGGCTGACCTTCTGTACTGTTTCC 60

DB 10274 GCCTGGAACCTGCACCTGCTCACCAGCGAGCTCCGCTGATTCATTTGCTCTTCC 10215

QY 61 TCGGCGTGGGCGCTCTTGGCTCGATCGCCAGGCGCACTA 101

DB 10214 TCCGCTGCAGACTGTGGGCGCGCCGCCAGGGAACCTA 10174

RESULT 10

AA168064 standard; DNA; 552 BP.

XX AA168064;

DT 13-MAR-2002 (first entry)

DE 9-keto-epothilone PKS inactive KR domain nucleotide sequence.

KW Cystobacterales; recombinant; polyketide synthase; PKS; polyketide;
KM antitumor; p11a gene; cytosolic; antiproliferative; antitumor;
KW antitumor; antitumor; antitumor; antitumor; antitumor; ds.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 1..552
FT /tag= a
FT /note= "mutated KR domain of extender module 6"

XX WO200183800-A2.

XX 08-NOV-2001.

XX 26-APR-2001; 2001WO-US13793.

XX 28-APR-2001; 2000US-0560367.

XX 14-SEP-2000; 2000US-232696P.

XX 21-DEC-2000; 2000US-257517P.

XX 03-APR-2001; 2001US-0825856.

XX 03-APR-2001; 2001US-0825876.

XX 13-APR-2001; 2001US-269020P.

XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Arslanian R., Ashley G., Frykman S., Julien B., Katz L., Khosla C;

XX Lau J., Licardi R., Regentin R., Santl D., Tang L;

XX WPI; 2002-075167/10.

XX P-PSDB; AB807168.

XX Recombinant host cells useful for producing polyketides e.g. epothilone

XX or its derivatives, comprises a recombinant expression vector encoding

XX a heterologous polyketide synthase gene

XX Example 11; Page 165-166; 221pp; English.

CC synthesized by the PKS enzyme encoded on the vector. An epothilone
CC derivative of a specified formula can be produced by culturing the host
CC cell with a dikeide equivalent compound of a specified formula. The host
CC cells produces polyketides at high levels and are used in the production
CC of not only epothilones, including new epothilone derivatives, but also
CC other polyketides. Methods of purifying the epothilone derivatives are

CC also useful for treating cancer, hyperproliferative diseases and
CC conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC /or stenosis. It improves polyketide production in any organism and
CC also for production of products of recombinant PKS genes and modification
CC enzymes. The present sequence represents the nucleotide sequence of a
CC mutated and inactive KR domain of extender module 6 of the novel 9-keto-
CC epothilone PKS of the present invention.

XX Sequence 552 BP; 70 A; 177 C; 212 G; 93 T; 0 other;

Query Match 45.1%; Score 45.6; DB 24; Length 552;
Best Local Similarity 66.0%; Pred. No. 0.00016;
Matches 66; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGCGGCTGACCTTCTGTACTGTTTCC 60

DB 349 GCCTTGAACCTGCACCTGCTCACCAGCGAGCTCCGCTGATTCATTTGCTCTTCC 408

QY 61 TCGGCGTGGGCGCTCTTGGCTCGATCGCCAGGCGCACT 100

DB 409 TCGGAGAGAGGCGCTTGGGCTCGCGCGGCGGCGGCACT 448

RESULT 11

AAV21186 standard; DNA; 5676 BP.

XX AAV21186;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.

KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

KM polyketide synthase; actinomycete; ansamycin; ds.

XX Amycolatopsis mediterranei.

XX Key Location/Qualifiers

FT CDS 3..5676
FT /tag= a
FT /product= "polyketide synthase"
FT /note= "no stop codon given"

XX WO9807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997; 97WO-EP04495.

XX 20-AUG-1996; 96EP-0810551.

XX (NOVS) NOVARTIS AG.

XX Engel N., Schupp T., Toupet C;

XX WPI; 1998-169172/15.

XX P-PSDB; AAW52844.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used

XX to produce rifamycin and rifamycin analogues

XX Claim 4; Page 38-43; 205pp; English.

XX The present sequence represents Amycolatopsis mediterranei strain wt3136

XX 5.7 kb KpnI fragment DNA, from the present invention. The present

CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC gene cluster DNA fragment comprising a DNA region involved directly or
CC indirectly in the gene cluster responsible for rifamycin synthesis,
CC including the adjacent DNA regions to the right and left which, by
CC reason of their function in connection with rifamycin biosynthesis,
CC qualify as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

CC Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

Query Match 45.1%; Score 45.6; DB 19; Length 5676;

Best Local Similarity 68.5%; Pred. No. 0.00027;

Matches 63; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 10 CTGCACACCCCTTACCCGAGCAGCCGCTGACCTCTGCTGTTTCTCGGCGTGG 69
DB 2631 CTGCACGACGTACCCGCAATGATGATCTGCGGCTGCTGTTCTCTCCGCTCC 2690

OY 70 GCGGCTCTGCGCTGCATGCGCCAGGCGACACTA 101
DB 2691 GGGGCTTCTGCGCTCCGCGGCGAGGCAACTA 2722

RESULT 12

AD17185 standard; DNA: 27541 BP.

AC AD17185;

DT 29-NOV-2001 (first entry)

DE Streptomycetes noursei nys2 DNA of nystatin PKS gene cluster.

KM Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; nys2; ds.

OS Streptomycetes noursei.

FH Key Location/Qualifiers
FT CDS complement (454..1191)

FT /*tag= a "NysF protein"

FT /note= "CDS does not include start codon"

FT complement (1275..3092)

FT /*tag= b

FT /product= "NysG protein"

FT complement (3070..4824)

FT /*tag= c

FT /note= "CDS does not include start codon"

FT /product= "NysH protein"

FT 5122..6156

FT /*tag= d

FT /product= "NysD3 protein"

FT 6338..27541

FT /*tag= e

FT /product= "NysI partial protein"

FT /note= "CDS does not include stop codon"

XX MO200159126-A2.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-GB00509.

XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SMTF) SINTER STIEMPELSEN IND TEK FORSK.

PA (ALPH-) ALPHARMA AS.

PA (SINV-) SINVENT AS.

PA (DZTE/) DZTEGLESKA H.

PA (ZORC/) ZOTCHEV S. B.

PA (SEKU/) SEKUROVA O. N.

PA (FJAE/) FJAEVYK E.

PA (BRAU/) BRAUTASET T.

PA (STRO/) STROM A R.

PI Zolchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI: 2001-557614/62.

DR P-PEDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.

XX New nystatin polyketide synthase polynucleotides and polypeptides,

PT useful as antibiotics and antifungals -

PS Claim 2; Page 151-166; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene

CC cluster encoding a modular type I polyketide synthase (PKS) enzyme

CC involved in the biosynthesis of the macrolide antibiotic nystatin.

CC The nystatin PKS is useful as antifungal antibiotics. The present

CC sequence is a Streptomycetes noursei nys2 DNA of nystatin PKS gene cluster.

XX Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

SQ Query Match 45.1%; Score 45.6; DB 22; Length 27541;

Best Local Similarity 68.5%; Pred. No. 0.00027;

Matches 63; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 10 CTGCACACCCCTTACCCGAGCAGCCGCTGACCTCTGCTGTTTCTCGGCGTGG 69

DB 24890 CTGCACGACGTACCCGCAATGATGATCTGCGGCTGCTGTTCTCTCCGCTCC 24949

OY 70 GCGGCTCTGCGCTGCATGCGCCAGGCGACACTA 101

DB 24950 GGGGCTTCTGCGCTCCGCGGCGAGGCAACTA 24981

RESULT 13

AAV21187 standard; DNA: 53789 BP.

AC AAV21187;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

KM Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX polyketide synthase; actinomycete; ansamycin; ds.

OS Amycolatopsis mediterranei.

FH Key Location/Qualifiers
FT CDS 1825..15543

FT /*tag= a

FT /label= ORF_A

FT /product= "polyketide synthase"

FT 15550..30759

FT /*tag= b

FT /label= ORF_B

FT /product= "polyketide synthase"

FT 30895..36060

```
FT      /*tag- C
FT      /label- ORF_C
FT      /product- "polyketide synthase"
FT      36259..41325
FT      CDS
FT      /*tag- d
FT      /label- ORF-D
FT      /product- "polyketide synthase"
FT      41373..51614
FT      CDS
FT      /*tag- e
FT      /label- ORF-E
FT      /product- "polyketide synthase"
FT      51713..55293
FT      CDS
FT      /*tag- f
FT      /label- ORF-F
FT      /product- "polyketide synthase"
XX      WO9807866-A1.
XX
XX      26-FEB-1998.
XX
XX      18-AUG-1997; 97WO-EP04495.
XX
XX      20-AUG-1996; 96EP-0810551.
XX
XX      (NOVS ) NOVARTIS AG.
XX
XX      Engel N, Schupp T, Toupet C;
XX
XX      WPI: 1998-169172/15.
XX      P-PSDB: AAM52845-W52850.
XX
XX      Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX      to produce rifamycin and rifamycin analogues
XX
XX      Claim 4; Page 53-102; 205pp; English.
XX
XX      The present sequence represents a Amycolatopsis mediterranei rifamycin
XX      synthetase gene cluster DNA fragment: from the present invention. The
XX      DNA fragment comprises a DNA region involved directly or indirectly
XX      in the gene cluster responsible for: rifamycin synthesis, including
XX      the adjacent DNA regions to the right and left which, by reason of
XX      their function in connection with rifamycin biosynthesis, qualify
XX      as constituents of this rifamycin gene cluster, and functional
XX      fragments, derivatives or constituents of these. The Amycolatopsis
XX      mediterranei rifamycin synthetase gene cluster DNA fragment can be used
XX      for producing rifamycin, rifamycin analogues or precursors. It can also
XX      be used for inactivating or modifying genes involved in ansamycin or
XX      rifamycin biosynthesis. The DNA can be used for constructing mutant
XX      actinomycetes strains from which the natural rifamycin or ansamycin
XX      biosynthesis gene cluster has been partly or completely deleted. The
XX      DNA fragment can be used for assembling a library of polyketides.
XX      synthetases, which can be used for assembling a library of polyketides.
XX      A hybridisation probe of the invention can be used for identifying DNA
XX      fragments involved in the biosynthesis of ansamycins.
XX
XX      Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;
SQ
Query Match 45.1%; Score 45.6; DB 19; Length 53789;
Best Local Similarity 68.5%; Pred. No. 0.0003;
Matches 63; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 10 CTGACACACCTTACCCGAGACACCCGCTGACCTTCTGCTACTGTTTCTCGGCGTCG 69
DB 45723 CTGACAGCGTGACCGCGCAAGTGAATCTTGGCGCTTCTGCTCTCTCGGCTCC 45782
QY 70 GCGGCTCTGCGCTCGATCGCGCAGGCGACTA 101
DB 45783 GGGGCTTTCGCTCCGCGGCGGCAAGGCACTA 45814
RESULT 14
AAF31641
ID AAF31641 standard; DNA: 6333 BP.
```

```
XX      AAF31641;
XX
XX      09-APR-2001 (first entry)
XX
XX      Mycobacterium tuberculosis mas gene.
XX
XX      Mycobacterium tuberculosis; attenuated microorganism;
XX      signature tagged transposon mutant; mutant library;
XX      mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
XX      vaccine; mas; mycobactericidal acid synthase; ds.
XX
XX      Mycobacterium tuberculosis.
XX
XX      WO200102555-A1.
XX
XX      11-JAN-2001.
XX
XX      06-JUL-2000; 2000MO-IB00950.
XX
XX      06-JUL-1999; 99US-0142982.
XX      08-JUL-1999; 99US-0142833.
XX
XX      (INSP ) INST PASTEUR.
XX
XX      Glacquel B, Guillhot C, Camacho L;
XX
XX      WPI: 2001-091804/10.
XX      P-PSDB: AAB66471.
XX
XX      Screening a mutant library for mutants unable to grow under specific
XX      conditions and for identifying loci involved in pathogenicity.
XX      comprises using signature tagged transposon mutagenesis -
XX
XX      Example 8; Page 147-148; 159pp; English.
XX
XX      The present sequence is given in a specification relating to a method for
XX      screening a library of mutants. The method comprises constructing a
XX      library with insertions in genes and/or regulatory regions of the
XX      organisms of interest, where the insertion contains a tag and/or a
XX      transposon associated with a tag. The mutants are identified by
XX      hybridisation of the tags to known sequences. The method is useful for
XX      treating an individual suffering from a mycobacterial infection,
XX      suspected of being infected with a Mycobacterium, or having been
XX      exposed to an infectious Mycobacterium. It is also useful for
XX      identifying and isolating mutants of actinomycetales and for identifying
XX      compounds that have antibiotic activity. The method is used to identify
XX      mutants of microorganisms, preferably an actinomycetales, such as
XX      M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
XX      M. paratuberculosis, that is unable to grow under specific conditions.
XX      It is especially useful for identifying loci involved in pathogenicity.
XX      It is useful in constructing vaccines. The method can be used to screen
XX      multiple libraries concurrently. It can screen libraries of different
XX      organisms or different strains of the same organism.
XX
XX      Sequence 6333 BP; 1010 A; 2010 C; 2202 G; 1111 T; 0 other;
SQ
Query Match 43.6%; Score 44; DB 22; Length 6333;
Best Local Similarity 65.0%; Pred. No. 0.00067;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 2 CATGGGTCTGCACACCTTACCCGAGACGCGCTGACCTTCTGCTACTGTTTCCG 61
DB 5636 CTGGAACCTACACACCGGCGACCCCTGGTACGCGGCTGCTCTGCTTCTCCG 5695
QY 62 CGGCGTGGGCGTCTTCGCTCGATGCGGCGGCACTA 101
DB 5696 CGGGAGCGGCGATTCGCTCGCGCGGTCAGGCGGCGCTA 5735
RESULT 15
AAF90035
ID AAF90035 standard; DNA: 8301 BP.
```



```

XX AAF90035;
AC
XX
XX
DT 06-AUG-2001 (first entry)
XX
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic;
KM type I polyketide synthase; ss.
XX
XX OS
XX
XX Key Location/Qualifiers
FH 1.8301
FT CDS /tag= a
FT /product= "type I polyketide synthase"
XX
XX WO200140497-A2.
XX
XX PD 07-JUN-2001.
XX
XX PF 27-NOV-2000; 2000WO-FR03311.
XX
XX PR 29-NOV-1999; 99FR-0015032.
XX PR 07-JUN-2000; 2000US-0209800.
XX
XX PA (AVET ) AVENTIS PHARMA SA.
XX
XX PI Jeannin P, Pernodet J, Guerin M, Simonet P, Courtols S;
PI Cappellano C, Franco F, Raynal A, Ball M, Sezonov G, Tuphale K;
PI Frostegard A;
XX
XX DR WPI: 2001-374849/39.
XX P-PSDB: AAB83972.
XX
XX PF Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics
XX
XX PS Claim 35; Page 302-305; 356pp; French.
XX
XX CC The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthetase) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. AAF90034-39
XX represent open reading frames (ORFs) of the coding strand of cosmid
XX a26g1, and encode type I polyketide synthases.
XX
XX SQ Sequence 8301 BP; 1426 A; 2629 C; 2669 G; 1577 T; 0 other:
XX
XX Query Match 43.6%; Score 44; DB 22; Length 8301;
XX Best Local Similarity 67.4%; Pred. No. 0.0007;
XX Matches 62; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
XX
XX QY 10 CTGCACACCTTACCGCGAGCGGCGGACCTTGTCTGCTTCTTCGCGGCTCG 69
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 7549 CTCACATCATCACACCGCGGCGGCTGCGGCGGACTTCTTTCATCGGATCC 7608
XX
XX QY 70 GCGCTCTTCGCGCTCGATCGCGGCGGCGGCGGCGGCTA 101
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 7609 TCGCTCTTAGTGTCTGCGCGGCGGCGGCGGCGGCTA 7640
XX

```

Search completed: November 5, 2002, 12:35:35
 Job time : 236.215 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds

(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050

Perfect score: 101

Sequence: 1 gcatgggtgctgcacacct.....tcgacgcgcagcgagcagcta 101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: GenEmbl:*
2: gb_ba:*
3: gb_htg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_rto:*
12: gb_sis:*
13: gb_sy:*
14: gb_un:*
15: gb_vl:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF217189	AF217189 Sorangium
2	101	100.0	68750	6	AF210843	AF210843 Sorangium
3	101	100.0	68750	6	AR193029	AR193029 Sequence
4	101	100.0	68750	6	AR199551	AR199551 Sequence
5	101	100.0	68750	6	AR199559	AR199559 Sequence
6	101	100.0	68750	6	AR199567	AR199567 Sequence
7	101	100.0	68750	6	AR201097	AR201097 Sequence
8	101	100.0	68750	6	AR208671	AR208671 Sequence
9	101	100.0	71989	6	AR172664	AR172664 Sequence
10	50	49.5	49736	1	AF319998	AF319998 Stigmatel
11	49.8	49.3	113193	1	AF357202	AF357202 Streptomy
12	48.2	47.7	16124	6	AX024384	AX024384 Sequence
13	48.2	47.7	123580	1	AP263912	AP263912 Streptomy
14	48.2	47.7	123501	6	AX211739	AX211739 Sequence
15	48.2	46.7	82746	1	AF453501	AF453501 Actinosyn
16	47.2	46.7	552	6	AX403009	AX403009 Sequence
17	46.6	46.1	6462	6	AX153794	AX153794 Sequence
18	46.6	46.1	34071	6	AX153790	AX153790 Sequence
19	46.6	46.1	42717	6	AX153789	AX153789 Sequence
20	46.6	46.1	552	6	AX403011	AX403011 Sequence
21	45.6	45.1	5676	6	A69718	A69718 Sequence 1
22	45.6	45.1	27522	1	AB070942	AB070942 Streptomy
23	45.6	45.1	27541	6	AX211706	AX211706 Sequence
24	45.6	45.1	53784	1	AMM223012	AMM223012 Amycolato
25	45.6	45.1	53789	6	A69720	A69720 Sequence 3
26	45.6	45.1	84985	6	SNA278573	AJ278573 Streptomy
27	45.6	45.1	90445	1	AF040570	AF040570 Amycolato
28	45.6	45.1	30000	6	AX250262	AX250262 Sequence
29	45	44.6	66808	1	SAU421825	AJ421825 Stigmatel
30	45	44.6	104326	1	AB070940	AB070940 Streptomy
31	44.6	44.2	10910	6	AX024383	AX024383 Sequence
32	44.6	44.2	10910	6	AX024276	AX024276 Sequence
33	44.6	44.2	6333	6	AX069081	AX069081 Sequence
34	44.6	44.2	8301	6	AX153792	AX153792 Sequence
35	44	43.6	9699	1	MSGWACSTN	M95808 Mycobacteri
36	44	43.6	14326	1	AE007123	AE007123 Mycobacte
37	44	43.6	24292	1	MTCY2461	Z83858 Mycobacteri
38	44	43.6	20394	6	SNA132222	AJ132222 Streptomy
39	43.4	43.0	20394	6	AX067996	AX067996 Sequence
40	43.4	43.0	30690	6	E38020	E38020 Avermectin
41	43.4	43.0	32870	1	AF007101	AF007101 Streptomy
42	43.4	43.0	64957	1	AB032367	AB032367 Streptomy
43	43.4	43.0	65140	6	AX211705	AX211705 Sequence
44	43.4	43.0	67523	1	SC024241	U24241 Sorangium C
45	43.4	43.0	67523	1	SC024241	U24241 Sorangium C

ALIGNMENTS

```
RESULT 1
AF217189
LOCUS
DEFINITION
AF217189 58733 bp DNA linear BCT 09-JUN-2000
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189
KEYWORDS
AF217189.1 GI:7453554
SOURCE
Polyangium cellulosum.
ORGANISM
Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 58733)
```

Pred. No. is the number of results predicted by chance to have a

QSTRMLMETTLDLKRRLPMLGDHVGAVVFGAAYLEMAISSGALGDGPIQITDV
IAEALAFAGDAVILVQVVTTEQPSRGLQFOLISRAPGASHAFRHARCALLREBTE
VPAGLTLSAVIRLQASMPAAATVAETIMETLOVQPAEGTIELMRGEEALGRRLP
DAAGSAEYRLRHPALLDACFOVYVSLFAGGAGATVWPVPEVSSILNLOPSELMCHA
RYVNHGROTPDRGGLDFWVVDSSGAIVAEVSGLVQRIEVEDWLELMEP
AAVGAIAKVNAGRWMLLIGGGGIGLAAALRSLTEGGAIVHAAASNTSAGVRLALKAF
DQAPTAHVHLGSLDGGELDPGLGACGLADPRADVSDPALDILALVYGCVMYTV
QALMGAFPDAPRLMLTRGAQVAGVSVQAPLGLIGRVIATHEHADLRCAVYDLD
PIRDPGELGALLIELLADDAEAEVALRGERGVCAVIRVROSPETIRRGRESVPVDT
IRADSTIYVGGIGLGLSVAGMLAEKAGHLVLYVRSAGSAVEEUPAAVALAEAGAR
VVAADVADRAQRLERILREVTSGMPKRGVHAAGILDDGLMOQTPAFKFAVAPK
VOGALHLHALTRAPLSFEVLVYASGVGLSGPGQNTYAANFTLALAHRAQOGLPA
LSVDGLFAEVMAAABEDRGARLVYRSGMSLTDPGEGSALARLIESGVCGVMYTV
PLWAEVLPAAASSRMLSLVTAHRASAGGAPGADGLRLRLIAEAPASRSGLLEPLLR
AOISQVRLPECKTEVDAPITSIGMSLIMGELRRRIEMGLGITYPATILMTYPIYAA
LSGHLARECEAPVESPHTTADSVEIEMSODDLTOLIAKFKYALT

16134. .37907
/gene="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="WTRGPTAQNPNLKOAITIQRLEBRGLAQAELEIETPIAIV
GIGCRPGGADAPFAFWELDLERDAVOPLDRLMALVGAPEAVPHNAGLTTEPIDC
FDAEFGISPREARSLDPOHRLLEVAWEGELSDAGIIPKSIDSGRTGVVGAFTADYA
RTVARLPREEDAYSATGNMLSIAGRSLYTLGLOGCLTVDPAKSSSLVAIHLACRS
LRAGESDLAAGVSTLSPDMMEAAARTQALSPGRCPTFASANGVFRAGCGGILVY
LKRLSDAQRDGRITWALIRGSAINHDSRGTGTAAVNTLAOETVLAERALSAREAGV
DYVEHGTGTSIGDPIEVEALATYTPARSDOTRCVLAQVNTNISHLEAAVAGLIK
AALSLTHERIPENLNFRTLNPRIRLEGSALALATEPWPRTDRPFRAGVSGFGST
NAHVLEAPAVEALPMPAPERSAELLVSGSEGAIDQARLRHLDHMLPELIGDY
AFSLATTSAMTHRLVAVTSREGLLALSAVQOCTPAGARCIASSRSGKLALFT
GOGAOTPGMGRLCAAMPFAEDRCVTLFDRELDRLPREYMAEAGSARESLIDOT
AFTOPALFAVEALTLALMRSWGVEPELLVGSIGELVACVAGVSLDEGVAVLARG
RLMOGLSAGGAMVSLGAPAEVAAAVAHAAVSTAAVNGPEOVITAGYEOAVQVIAA
GPAANGVTRKRLVSHAFHSPMLPEPMLFEFGVAASVYIRRSVSLVSNLSGKVVYDTE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 5.5e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGGTCTGCACACCTTACCCGCGAGACAGCCGCTGACCTCTTCGTACTGTTC 60
|||||
DB 19950 GCATGGGTCTGCACACCTTACCCGCGAGACAGCCGCTGACCTCTTCGTACTGTTC 20009
|||||

QY 61 TCGGCGTCCGGCGCTCTTCGGCTCGATCGGCGCAGGCGACGTA 101
|||||
DB 20010 TCGGCGTCCGGCGCTCTTCGGCTCGATCGGCGCAGGCGACGTA 20050
|||||

RESULT 2
AF210843 68750 bp DNA linear B-T 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epithione biosynthesis gene
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
1 (bases 1 to 68750)
Molnar, I., Schnupp, T., Ono, M., Zirkle, R., Milnamow, M.,
Nowak-Thompson, B., Engel, N., Toupet, C., Straumann, A., Cyr, D. D.,
Goriach, J., Mayo, J. M., Hu, A., Goff, S., Schmitt, J., and Vigon, J. M.
The biosynthetic gene cluster for the microcubule-stabilizing
agents epithionones A and B from Sorangium cellulosum So ce90

JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar, I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source
CDS
1. .68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
<1. .1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSAARYLDDGHRGATERHVLARGLIEDLRALREH
LRIOGGPSEFHCMLGDLTVLELDLADQPLASTISFHHASLRHPDWTSDAMLYDGPALV
RWLAARGAAPLPREYEERERARTAOEARLMLAAPPCFAPDLPRFEDDANGPLGP
MSPEVAEARLRASVYARPELCALMLMGAGAPMSGYPAYEMLRPEMLILGFLPT
AIAASAPCTSEALRGARLPFASGEVVSSEKSGOIGNIPALMEBLRRTIVRAMGAND
LSRPERAIAIEVRLRQAPAPPAAGCLAVAGYSSSGRLSGLVTDGALISGCGND
IVMFQGRISPVLLAGTDPELPPLAUSOMLVAHANAQTISKVLTGESSPLIVARN
QARPSLVHARGFMAMVNOAMPDDEKCAPFVYVORSTIMEFHPPRCRLHEPAGSAFS
IACDESHLWCLSLAGRLLELMPHRRPGAPAFAYLGEHPYAAVYSLTANAHVL
MADPPRRLADKRGVETVLAETLRHPPHVAHVEDNRDIFALTQOPSDRWVHEIR
SGASVYVDYQOLMDRPMVLRNRRGLPFTTNDRLITLARS"
complement(1900..3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to D-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAAPCALGAESEEGOGDAGAGALAPAREVMAAEVA
AGOMGAVMLVARGDVHDVADVAGELGSAEMRDTFRISMTKATATVMMVLE
EKLIDLSVDVDRMLPELANRKVLARDGIDETVAERIPTRDMLTFMGISFEDA
SSPTORAIDELGLVNAQVPVPTPHGDEIRRLGLTLPMDHOGAMKNTGSLVGYLV
VGRADYQGFDAFVREIRLAPIGMDTDHFVPADKLARFAGCGYFTDEQTESTRMDRL
GAESAIVASPPAPPSGAAGLVSTVDYTLFARMIMNGVHEGRLLSASVREMTADHL
TPAOKAASEGFPFETHMGYGMVTAAPDAVSEVPGRYGDGSGFTSWINDPEREL
IGIWMOSAGFLFSGALERFMSVYVATESA"
3415. .5556
/note="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGTEROVLISLTALILTYTARSGELARLRKPEVLEGELFC
GVVLGPVVGALAPGFHRLAQEPAPVAVGLSISIGALLLLLMAGVIEADVGLKEEA
RGALSALGAIAPPLAAGAFSALVLDRLPGSLGLTGLSTAVASIAKVILIEBES
RRSYAOVTLAAGVSEVAAVNVAMTSSYASPALAVARSLASGLAFVILVYGR
LTHAMRWVADATRVSGOVSLVLTFLAAALRLQRLGHPILGAPLGGVILNSAPRT
NRPLDQVOTLVAGCFAPFVPLACMRDVSDLRPAMGTATALLATATATAKVPAA
LGARLGGKSGSEALVAVGLNKGGLDILVALIVGELGLSLEATYTNAAVALVTYTA
SPALLIWEKERAPPTOESARLERREARATVIVERLIVPVAHALPGFATDILVES
IVASRKRLGETVDITELIVEQDAPGPSRAAGSAGLARLGRVLRVLMWRQRELRS
IOAIRASDHDLVLVIGARSAPARAGMSGRLQDAIVORAESNVLVVVGDPAPARAS
ARLITVITIGETSYFAADIALVALAMDALVELILSSAOTDGAVWMBDRBSRVAV
ASVYDAVEAFRGRCGVRSVYHGARGSDPITRELARAPDILLVLCYDHGPTGLRL
YLGSTVESVYVNSRPVALLVAHGSTRQVCR
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11

<p> /product="nonribosomal peptide synthetase" /protein_id="AAF26925.1" /db_xref="GI:6724259" /translation="MTINQLNLELEHGCIKAADGGERLQIDAPKNAKLNPLNLAIRISEH KSTILLNLRPLAESIVAPAPERRHAPPLDIDQESTWLGRTGAFVPSGIHAYRED CTDLIDYPRLSRARFARVVAHDMIRAHPTLPMQVIEPKVDADIEITDLRGIDSTREA RVSILRDAHMRHLYDTERPPLVAVARLDEKQRLVSLIDLWDLGSLITFKML SFVEDPETSPLVELSYRDIYVALSRKSKSEAHORSOMYMKRRLAELPPPTPKMAD PSTLKEIRFRHPTESQYIPDSQMKRIKRVGCEGLPTQVYLAAEVSIVGRSASPRETL NITLENLPHPRVANDITGDTFTSWLDDITDTRDKSFQRAKRIQEDLRAMDHCVS GIEVQRAAAVLNIQIGGAFPVLTSLNQOYVGTSLQRLGTVPYSTQTPOLLDH QLEHDEDLVLANDIDVGDFPEVLAFAREQLPMQAAVSAKTLTEPEMGDEVORSLPQGLE ARASANTNLNLESEHMLVLAFAREQLPMQAAVSAKTLTEPEMGDEVORSLPQGLE EGRAPPTLVAIVMEKGMQVAVAVLALESGAVYPIDADLPARHLYLDDHEGVKL LTPQWMDGKSWPQGIQRLISEACVEEDGQPPMPPIQPSDIAVLYTSGSTGLY GVMIDHGAVALNITLDINERETGGQRYVALSSLSLTDVYDGLIAAGTIVYDPA SKLMDHGAVALNITLDINERETGGQRYVALSSLSLTDVYDGLIAAGTIVYDPA SKLMDHGAVALNITLDINERETGGQRYVALSSLSLTDVYDGLIAAGTIVYDPA EALPELOAIIPGVSYLSLGGAEASLWSIGYVRNVDLSNMSITPYGRPLRNQTFHYLD EALPEPRVAVGQLYIGGVALGAWDEKTRKSTLVHPETGRLTKTGDILRYLDP LIEPFMEKREDNQIKLGGYVLELGEITETLKSHPRNDVITVPCNDANKLLAAYVP EGRERRAAGDASLKERIDARHAAEADGSLDGRQVQKARRGLRDLDDKRPVYL TGDQPRRAGDADVARRSVRTLEAPITPEVGRGLSCLSVEPDGATLFRKFRPSAG SYTPVQYVAIGVAGKRGLEGVDEGGYLYHPEHMLKILDSHGLEBAHVRQNFVDEEA ENLFLVRIADAISLTYGSSRRECLYGAQMQLMEQAPSCNIGYQPOQNFEDQVR PVLDLRSDVYVGMGLGVRDPROVQCTIGDSSPRRATRGAPGRQHPADMLEI FLRLKLEPVYVPEVLEDALPLTNSGYDKRALPERKDTSPRSGHTAPDRALEI LVAAREVLELEVGLQGSFVDGATSHIYVMSLLOKRLDREIATLETFQYPNIGS LASGLRDSHDLDRPNMDDREVRKGRRS" 16251..21749 /gene="epob" 16251..21749 /gene="epob" /note="Epos B" /codon_start=1 /transl_table=11 /product="polyketide synthase" /protein_id="AAF26920.1" /db_xref="GI:6724254" </p>	<p> 100.0%; Score 101; DB 1; Length 68750; identity 100.0%; Pred. No. 5,4e-16; conservative 0; Mismatches 0; Indels 0; Gaps 0; </p>
--	---

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 60
|||||
Db 2562 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 25621

QY 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 101
|||||
Db 2562 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 25662

RESULT 4
ARI99551
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 60
|||||
Db 2562 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 25621

QY 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 101
|||||
Db 2562 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 25662

RESULT 5
ARI99559
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 60
|||||

Db 2562 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 25621
|||||

QY 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 101
|||||
Db 2562 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 25662

RESULT 6
ARI99567
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 60
|||||
Db 2562 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 25621

QY 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 101
|||||

RESULT 7
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 60
|||||
Db 2562 GCATGGGTGCTGCACACCTTACCCGCGAGACACCGCTGGACCTCTTCGACTGTTTCC 25621

QY 61 TCGGCGTCGGGCGCTCTTCGGCTCGATCGGCCAGGCAAGCTA 101
|||||

Db 25622 TCGGCGTGGGCGCTTCTGCGCTGCATCGGC/CAGGGCAGCTA 25662

RESULT 8
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6:03787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
Location/Qualifiers:
source 1..68750
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGTGTGCGACACCTTACCGCGGAC/CAGCGCTGAGCTCTTCTGTTTCC 60
|||||
Db 25662 GCATGGTGTGCGACACCTTACCGCGGAC/CAGCGCTGAGCTCTTCTGTTTCC 25621
|||||

QY 61 TCGGCGTGGGCGCTTCTGCGCTGCATCGGC/CAGGGCAGCTA 101
|||||
Db 25622 TCGGCGTGGGCGCTTCTGCGCTGCATCGGC/CAGGGCAGCTA 25662
|||||

RESULT 9
LOCUS AR172664 71981 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6:03342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 71981)
AUTHORS Julien,B., Katz,L., Khosla,C., and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
Location/Qualifiers:
source 1..71981
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71981;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCATGGTGTGCGACACCTTACCGCGGAC/CAGCGCTGAGCTCTTCTGTTTCC 60
|||||
Db 19950 GCATGGTGTGCGACACCTTACCGCGGAC/CAGCGCTGAGCTCTTCTGTTTCC 20009
|||||

QY 61 TCGGCGTGGGCGCTTCTGCGCTGCATCGGC/CAGGGCAGCTA 101
|||||
Db 20010 TCGGCGTGGGCGCTTCTGCGCTGCATCGGC/CAGGGCAGCTA 20050
|||||

RESULT 10
LOCUS AF319998 49736 bp DNA linear BCT 28-MAY-2001

DEFINITION Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
complete sequence.
ACCESSION AF319998
VERSION AF319998.1 GI:14210834
KEYWORDS
SOURCE Stigmatella aurantiaca.
ORGANISM Stigmatella aurantiaca
Bacteria; Proteobacteria; delta subdivision; Mycobacteria;
Mycococcales; Cyctobacterineae; Cyctobacteraceae; Stigmatella.
REFERENCE 1 (bases 1 to 49736)
AUTHORS Sllakowski,B., Nordstiek,G., Kunze,B., Blocker,H. and Muller,R.
TITLE Novel features in a combined polypeptide synthase/non-ribosomal
peptide synthetase: the myxalamid biosynthetic gene cluster of the
mycobacterium Stigmatella aurantiaca Sgal5
JOURNAL Chem. Biol. 8 (1), 59-69 (2001)
MEDLINE 21110452
PUBMED 11182319
REFERENCE 2 (bases 1 to 49736)
AUTHORS Sllakowski,B., Nordstiek,G., Blocker,H. and Mueller,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig
38124, Germany
Location/Qualifiers
FEATURES
source 1..49736
/organism="Stigmatella aurantiaca"
/db_xref="taxon:41"
complement(2..1306)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="putative aspartokinase"
/protein_id="AAK57182.1"
/db_xref="GI:14210835"
/translation="MRPISARAGRTSALALIYOKYGTSGVDETRIKYAKKRIA
ACKAGHDVVVYVMSGFTNRLLIKYSOTTEPNPEEDPVVATGQVSGICVALAIO
TQRRKAVSTFGHOVRIITTTSTSKARIKSIDERYDALAKKNIVYVAFQGDEN
VTTLGRGSDTTAVALAALAKADACEIYTDVGVYTTDNNVPAARKLERIAVEEMLE
LASVGAKEVLIQIRSEVFAMKRYKPLVWKSSTDDPGTLVCEEDPAAMDVYVSGIADKN
EAKIAIRGVDPVGVNAKIFGALDEQSIYDLIVONVSKDGRDVTFTVGKADLRKAK
DYKVIKASVKAEGVETDDQVSKSVIGVMNRHSGVAKMFTVLVAGAVNVQMISTS
EIKVSCVHSNTELAVRSLHRFAFGDKREAGAVSEVSLK"
1363..2895
/note="ORF2"
/codon_start=1
/transl_table=11
/product="putative histidine ammonium lyase"
/protein_id="AAK57183.1"
/db_xref="GI:14210836"
/translation="MSRPRNLIDGDTLKEILQVARTVVELAPAAARVKAARDL
VBRVNAAGDPSYGINFGTTLAEVRIDKRDRELQRLILSHAAGVSPLEPRAVL
LLRKNVLAKGSGIRPETLALALEMLNDVVPVPERGVSAGSDLAFLALAVFI
GGEAFYKGERLPAQALERAGIKPVLEKAGLALVNGTQAMCAVGTILOARMLA
DLADLAGAMTLEGLIGSHKPTPETODVVAHKGOKACAHLEILLADSLVSHVNC
KYODPSLAKMPOVHGAREGLSPARRILEVINSATDNPFLVFETERIVSGNTHQ
PVSLADVAAMALVDTLSAISERVEDLVNPALSGLPPELAKNSGLNSGPMIAQVISA
LVAESRVLSPASVDSIPSSAGREHVSGMTALAKGRQVADFTSCAIELLVAAQ
LDYROPTRAGKGPQAAVELIRSKIPTEMKDELRHDLIAVSAALIDSGELNVPATR
60"
complement(3275..7822)
/gene="mxaa"
complement(3275..7822)
/gene="mxaa"
/note="non-ribosomal peptide synthetase; contains a
condensing, adenylation, and PCP reductase domains"
/codon_start=1
/transl_table=11
/product="Mxaa"
/protein_id="AAK57184.1"
/db_xref="GI:14210837"
/translation="MSVRELTTTAAKGVRLRLEGGDKLOVEASKAVINPELPAEELAN
KADIIAFLAKHSSRETSGLQPLARPEVIDPLSSGQERLWMLDLSPETSQVYVHIG
LNAKGLDLTKRSLSDELVRRHEVLRHFFPEVSPROVIVSPUGVALTIVELPGL
TSSOREEELIYRSDELADQPFDLAKGPELRLVVALGPDDEFALFVTKHHLITDGWSLG

REFERENCE 1 (bases 1 to 123580)
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strim,A.R.,
Valla,S. and Zotchev,S.B.
TITLE Biosynthesis of the polyene antifungal antibiotic nystatin in
Streptomyces noursei ATCC 11455: analysis of the gene cluster and
deduction of the biosynthetic pathway
JOURNAL Chem. Biol. 7 (6), 395-403 (2000)
MEDLINE 20334850
PubMed 10873841
REFERENCE 2 (bases 1 to 123580)
AUTHORS Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
Valla,S. and Zotchev,S.B.
TITLE Direct Submission
JOURNAL Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
N-7489, Norway
FEATURES
source location/qualifiers
1..123580
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="taxon:11455"
/db_xref="taxon:1371"
complement(46..733)
/gene="nysF"
complement(46..733)
/gene="nysF"
complement(46..733)
/function="presumably post-translationally modifies ACP
domains on PKS"
/note="putative 4'-phosphopantetheine transferase"
/codon_start=1
/transl_table=11
/product="NysF"
/protein_id="AAE71762.1"
/db_xref="GI:8050836"
/translation="MIEILPATVATEAVDDRRPGRLLSSREVIATAVSROE
FTVRIHARRALRIGIHVDRAILPNRGAGPWPQVIGSWTHCAGYAAAVSAEISA
AVSIDAEPKPLPAGVLIATLPSERPHLYAAHRPDVMDRLSSAKSVKRAMLP
LTQRELDSEAEIVDPGCAFTARLLVPGPLLGGRRVTPGKWHSTPALTTAVHL
PAPTRDRHRHRLTVHSPLPRPTEG"
complement(867..1684)
/gene="nysG"
complement(867..1684)
/gene="nysG"
complement(867..1684)
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysG"
/protein_id="AAE71763.1"
/db_xref="GI:8050837"
/translation="MASFDLEEERTAPRRARLVGLRPHRSVALAVSMVGIVL
NAFGLLGRVLDLADGVGVGPAPGIDPAIGRLVLLVLAIVVASLPMIAQR
LVASAVRTHELRLRPAEKLRLPLRHPROPAGELSTRTNDINDLLOOTLAE
LTSISLRTLMVLMVLSPLAVVMLTSPVSAIIAARISKRAOPHYAOKANSATL
NAHVEVCTGHALIKGFCRRAAEERPDACNDVYRAAQAQFASGAMPEVMEVAVL
GYLVAVTGMKVINGTITLGDVQAFLVYRQSPQVETIASVAGRLQSGIASAORF
TLDDAEQADPLRPGTPARAEGRVETDVSFSPDPLIENMLSLVEGSGTVAIG
PTGAGKTTGLNLMRFYEPDPSGRLDGTATMTDRDLRSRGVIVADDMVGGTIA
ENTAYAGPACRADIEPARATGACDRITRLPGCYDVLDDSGTSAAGKOLLTVAR
AFLARAVLVLDATSSVDTREVLTORANSLRAGTSTFVIHRUSTINDADLIYVM
DAGRIVEGTHDULCAQGLYARLHAARTHTPTGAAG"
complement(2662..4416)
/gene="nysH"
complement(2662..4416)
/gene="nysH"
complement(2662..4416)
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysH"
/protein_id="AAE71764.1"
/db_xref="GI:8050838"
/translation="MLRLRQAQLRPYAMATAVLAQLVQILGTLPLTGLAALIDQ
GVVRDGGRTTELVGMVVALVOIAALGAALAAATATATAMGRDLRSALFRILDS

ARRIGRGTPELSTRVNDVQVONLAQTEGFCIVCAPMLCISGLLALADQVPLAL
LVALLVAVACFCILLARMTLARMQTLIDRGLRLREITGVRRVRSVDRDHRRA
RFAQNDAPFLVRSRRGRRLATMLPVVLLMDPEFTVALRSHRIDAGMPGSGISA
LSYLSITLMSVYMLAFVFLSVRAVQAGRIAEVDTGSSVPAAPQVRBPARI
ELCAGGRYGAEPVLRVDVLTVEGERGAVYAGSGKTTLLNLTIRADYTBAY
RVGCTGRRELTAAVAVGFVQRPRLVAVSMDLRFRPATDEELVRAVQA
ADVARMPDGLDAEITOGGNGVSGQORSLRALLRREIYLFDDCSFALDQADA
ALRTVALPYTAGATVITVQORISAGHDADRIVLDRGVAAQCTHVVLLRTSPYREI
ALSQLTSEEAHGLAGR"
4714..5748
/gene="nysDIII"
4714..5748
/gene="nysDIII"
/function="presumably involved in mycosamine biosynthesis"
/note="putative dgp-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NysDIII"
/protein_id="AAE71765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITGODGSYLAENLLSOGYVWGLIRGOANPKRSRV
SLASLDFITDGDMDOGSLVAVDVQDEYVNLGAISSPMQOAEIYMEVNGMG
VLMLAIRVNSLSTSRYSRPGQIRFQASSEKPKRAEIPQETTLFHRPSYG
AAAYGHTYTRNTRESGMYAVGMLFNHSPRRGEFYTRKISLAVARIQGLDKL
ALGNLDVARDMGVAGDYVRAMHMLQDADDDVYIGTGMHNSVRAVRIAEFHVGLNW
EDYVVIDPDLVRAEVEVLCAQSAKADRIKMPDVDFPLMRMWDSDLAQVRENO
YGVLLAANW"
5930..34363
/gene="nysI"
5930..34363
/gene="nysI"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="NysI"
/protein_id="AAE71766.1"
/db_xref="GI:8050840"
/translation="MNEOKLRDYLKLTADLRTRRRVYKLESAAQEPVAILGMC
YRGCVSPEDIMRVNVEAGHGVNPPEDQMDLEALAAPRTASGFLVAPDPAEDP
GISPREVAMPDQQRVYLESAMEAFERAGIDPPSVKSGRTGFTGAMADYRGPADG
AGEFQLTNGTGVSLGRISTYFTGVPAVYVDVACSSLVAAHLATQALRAGCTIAL
AGCVTITNSGPGTIEMGROGTSADRCRSFGDTADGTGMAEGVGLTVERSDAVRN
GHEILAVRGTAANOGASNGLTAPNGSOQVITQOALVNAALADIDIVEAHGCT
TIGDVPVBAQALLATYGGONRPAIDRPLIGYSKNSLSHTOAAAGVAYIKVMVMBGTL
PRLHAEPHNHNDMSOGARLLTPTDTPMTGAPRAAVSSGICSTNAHTITIBAP
EPQEDNATIQDDAAGSTPATVAVPVGVVLSGTPOLKQQAALRAIDTGRPR
DLDLNLSIATTRAGFENRAVLLATDHPALTDGLTLAADDPAAAPAVITGTRAT
RLAVLFTGQAQRLGAGRELAARFPAFAALDADLAFPHDRPLREVLMDADALL
DRAVYKOPALFVEVALYRLIESGVRPHLDGSHGETVIAHLIAGVLSLADATLVA
ARGRLQALDKGAMTAVOASEADVAFLAGHEDOVAIAGVSPASVAGSEATVA
LAQILAADCKTRRLVSHAFHSPDLMEPLHDAFRAYEDULTDPLLPVYSNLTKPA
TVAQULTSADVWDVHRAVAFADGIDMLARHDTAFLEIGPQVLSAMQDCIDADA
DAVTLALRGRREERTLTTLALNHGATLDMTCFAGTGRDRDPLTYAFORRY
WPAULSGRLDSVSLGAAHNLHLSAAVSLDAGSTLTGRLSRQTHMLADHIVRG
TTLPTPGATLHVRAGDEYGCORVELTIAVLLPBGQVGVOLNPNVDSGRT
VYHNAEPDQDPTPWTNATGVTITTDASRQLPASEGCGTLPAGDHPALDAQMP
AGAEPLPDSGHTRLADGGFGTGPVTOGLEBAHMGDDVYAEIPEAGRSDEAREGL
HPALLDAHLHAAPFTGLSEGRGLPFWMEGVSLLHAGATVTLKVRILTAPVADALATV
ADTGAPVLSVSLVLRSVATQQLDTPAAVARDALPDLPTVPQATPBGVALALGA
DPGLTNGHGFADAPVAPVPLAALAADGPTTVVLSLADGDDADPARSARCABE
ALAAVOTWLDHHERFAARLVFVTRGATVGRDVAANAANGIVRASQSENGCFALVD
DPDGAIGAAALVVALYSGEQALAVRDVLRVARTVRRPLTEVGAAGDGTGCGDSSG
VPSGGAIVLVTTGGTGLAVLARKHLVAEYGVDDLVSRSGRAGVAGELVAEILAG
GARVRVACDVDDRAAVVELVGGHVASAVHAAGVLDGDMVGLTERISAVLRPVVD
AAVHLHEATRGDLDLDAFVYESSLAGVFGSGQVYAAANAFLDALMTRRRAREPLDS
LAWGPEWOSGSGTGLTDVDAERLARSQPSLVAGLFLDAVAGTATGATCVPVLD
LPVLRARGVPEPLRLSLIRVARRAAVAGSARTVLAQRLRDEGRDEMVDIVRG
OVAVLYGHATGCDVDAGRAPRDGSPSLAVELRNLTNTVGLRDLPATVLPYPTVYRH
LATVYDELDELGTDAEVATVQPAVVAADPTVTVGACATPGGVSSPELMDRLTEGT
DAVSGPTNGMDVESLJHPDDPGTSTYRSGCFJHAGEEDPDGFGSAPRALTD
SOORLLESWEAIERAGIDPVSLSRFGVAGVMSYSLMASLSPPEFGCGSGSS

PSLSGRVATYLTGLEPAPVTVDTACSSSLVAMHMAQALRSCEGLALAGVTVNSTP
AVDFPARGRCIPDGRCKAFADADAGVMSGVCVLYLVERSDVLRNGHELTAAGA
SAVNDQASNGCLTAPNGPSOORVITROALASGLTIGDGVVVAHKTGTGPTTAAO
LLATYGRDREPERELLISVYSNGLHTOAAAGVAYIKVYLAIRGVYPRTHVAPS
SHVDSSEAVELLSQOAMPETGRVRRAGVSFSGTINHVITVQAPGAKITAAAG
ARPTGAPVAVLLSGRARSALRGOARLLGHLOARDAELVHALATLRTSRFEEDAA
VVAODRDLIASLIGLADRPDPAYEEAAGRGRTAVLFTQSGORAMGSELEHVO
PEFAAPFAPVACFDPDLDRPLREYVFEAEAGSDSEALDETMTQOPALFVAVLEFV
VESMGVRDPVAGHSIGELIAAHVAGVITDEPACVLAATFIMOTALPRTGMAITOA
TEDETAHLDDTVAAIAVNGPSVYISDDEBAETIAATFPERKTKTRLRVSHAFHS
PRMDMDLAFRIVAEGLYRAVRIPLVSDLTGRDDDEVCETAEYVWHVRVAFAD
CVRLTRDAGATFLELIGSDGLTMAAEDTLGDHDAELVPMALRAGABELAAATLAR
LOVRGVVDMAVYLAGTGARRTDPTPTAYFOHAYWPOLPPTPALAALADPAOQJLMAA
VERGDARELADIIGLEBODLPLDLSLPAIWSRGNOKHLLDRLRVRERTRSKP
TAPVLDGTWILVADATADOPALLDGLASGABRNOEKHLLDRLADREDFVLRGI
ARTADVADATOVLSYLPIDERRADDCPPLTRALITVVALVQALATGACGRMTATRG
AVSTIPADPVTHPVQAAAMGLRGVALHPRVGLVLDPOVEDRAGORLGLIAYK
DAPDEDDVALRATSGRRLVRHTVEALPTAERTAGTVLTGTGGGLGEVAVRWL
ARAGQHLVLTSRGPDPAGAAELRAELEGGPSVAVACVADADALAAVLTALPEE
LPLTGVHTAGVGHGPIDITLSTAEFAGITAKLAGAHLIDLADREDFVLRGI
AGVMSGNSAVGANAYLIDALALHRRAGLAATSVANGPAAEMADDAVSELR
OGLGLDPAAPMTELRVAVRODVTYVADVMORIALFTVSARSALTAGLPEERAF
AADERTEDATGASEVTVRRLAEPEDRLTDLVRESATVLSHSSADAVEGRAL
RDVGFDSLFAVELRKRIGATGLSLPTMVEPDPPLLEAOLYLRKEIIGAVLEVAGPV
ATCGADDEPIAIIIGMACRFPGVSSEPOLMDLVAGTDAISEFPVNRBMQTHLEDPD
PDRRTTSTOGGFLHEDDEPDPFPGISPRALVMDPOORLLITMTESFEERACITP
ETLRSTLTGTGVSGYCEGACGAGDGTGSHWYTGSPSYLSRSLSYVGLCEPATYD
TACSSSLVALHACOSLRKNSNLVAGAGATTMTIPNPFILASRRLAKLADKCRAFS
DDADGMTLAEGVAVLVERLSDAQNRNGHVRHLSAINDGASNGLTAPNGPSQOR
VIRQALANARLAPGDIDALEAHGTGTPLGDIPTAOLATYGRDNDPSALLGSVKS
NIGHTQSAAGIASVTKMYMALRHSLEPTLHADASSHVDMGAGVRLTQARMPET
GRPRRAVVSFGISGTNAHVILLEQAPVADTPEEPVAVAPVITAGVVPWYVTASAA
ALRQGERLIAHAETVGTALPAPGIDIGLSYSKARPERHAAVVPYPRAGTPIALR
AVATDDESPVYARGVADVEGRVYFPCGSGWVGMSGLADESAVFAERLAECAAL
AEFTDMSLVDLRGVYAPSLERVDVQPSAFVAVSIAALMRSRGVLPDAVVGSHQ
EIAAIVVSGALSLRDGARVALRQALRAGLRGMSVALSYDLEPRLVEFGRV
SVAAVNGPRSYVAGEPALDALHARLTADIRARRIANDYASHSHOVEDLHEELLEV
LAELAPRTSEVFEFTVGTMDLITRMDAGVWFRNLRGVRPDAADLILAEYRAVY
ESSHSPVSMVVOEAIDEAGVPAVAGTLRDOGTDRFILSAEVPFRGVYDWAAGL
FEGTASRIDIPTTAFQHEHMAVYPAPBAYAAADPPDAETVAVEDGVSALTALAG
TDEDSVAALVLPALTSWRRARNDSTVDAMRYRVANKPLGGTLPHPSLTGLTWLVTADG
IDDDVAGALEFTYGAVERRLVLEDECVDRAVLRELRAGAEDVTGTSVYLAALERTDAG
PGTSLVLTALTALVALIQALGDAEIDAPVALTRGAVSTGRADLETPAVOAYGTGIGMT
AALHPQRMWGTLDLPAALDARAQRILAAVLSGALGSDQLAIRPSGVFTRIYABA

Query Match 47.7%; Score 48.2; DB 1; Length 123580;
Best Local Similarity 67.3%; Pred. No. 0.013;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 GCATGGCTGCTGCACACCTTACCCGCGAGAGCCGCTGACCTCTTCGTACTGTTTCC 60
Db 55337 GCCTGGCATCTGCACGCCACCGCGCTGCACCTGCGCGCTTCGTCTACTCC 55396

QY 61 TCGGCGTCGCGGCTCTTCGCTCGATCGGCGAGGCGAGCTA 101
Db 55397 TCGGCTCCGCGGCTCTTCGCGAGCCCGCGAGGCACTA 55437

RESULT 15
AX211739 125401 bp DNA linear PAT 06-SEP-2001
LOCUS AX211739
DEFINITION Sequence 35 from Patent W00159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS
ORGANISM Streptomyces noursei.
SOURCE Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 125401)
Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sielta,H.V. and
Gulliksen,O.M.

TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNIISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; SINVENT AS (NO) ; ZOTCHEV, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Stroem, Arne Reidar (NO) ; Valla, Svein
(NO)

FEATURES
source location/Qualifiers
1..125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"

BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN

Query Match 47.7%; Score 48.2; DB 6; Length 125401;
Best Local Similarity 67.3%; Pred. No. 0.012;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 GCATGGCTGCTGCACACCTTACCCGCGAGAGCCGCTGACCTCTTCGTACTGTTTCC 60
Db 55745 GCCTGGCATCTGCACGCCACCGCGCTGCACCTGCGCGCTTCGTCTACTCC 55804

QY 61 TCGGCGTCGCGGCTCTTCGCTCGATCGGCGAGGCGAGCTA 101
Db 55805 TCGGCTCCGCGCTCTTCGCGAGCCCGCGCGAGGCACTA 55845

Search completed: November 5, 2002, 12:11:44
Job time : 667.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_18300_18400

Perfect score: 101
Sequence: 1 ccgagcagctgctgctgcgc.....aagcgctgcctgctgcga 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	40.2	39.8	512	10	AM163063	au91g11.y
2	38.8	38.4	1156	14	BQ922705	BQ922705 AGENCOURT
3	37.6	37.2	442	10	AM620463	AM620463 sj05e03.y
4	37.2	36.8	477	12	BF598970	BF598970 sv24b02.y
5	36.8	36.4	512	9	AU183888	AU183888 AU183888
6	36.6	36.2	1025	14	BQ926826	BQ926826 AGENCOURT

Result No.	Score	Match	Length	DB	ID	Description
7	36.6	36.2	1065	14	BQ669726	BQ669726 AGENCOURT
8	36.4	36.0	462	9	AU081551	AU081551 AGENCOURT
9	36.4	36.0	1389	11	AY105590	AY105590 Zee mays
10	36.2	35.8	1145	14	W29156	W29156 mb96a09.r1
11	36.2	35.6	420	10	BE588131	BE588131 WHE0657.G
12	36.2	35.6	483	13	BQ254408	BQ254408 BQ254408
13	35.6	35.2	554	17	AQ427466	AQ427466 CITR1-E1
14	35.6	35.2	596	13	BM656407	BM656407 170006873
15	35.6	35.2	611	13	BM615342	BM615342 170006871
16	35.6	35.2	613	13	BM605012	BM605012 170006870
17	35.6	35.2	617	13	BM615807	BM615807 170006592
18	35.6	35.2	632	13	BM601596	BM601596 170006870
19	35.6	35.2	640	13	BM578146	BM578146 170006874
20	35.6	35.2	671	13	BM584980	BM584980 170006872
21	35.6	35.2	641	13	BM597907	BM597907 170006875
22	35.6	35.2	688	13	BM596795	BM596795 170006874
23	35.6	35.2	689	13	BM645847	BM645847 170006873
24	35.6	35.2	694	13	BM597110	BM597110 170006874
25	35.6	35.2	764	17	AQ327293	AQ327293 nbxb0040P
26	35.6	35.2	1255	14	BQ937459	BQ937459 AGENCOURT
27	35.4	35.0	385	13	BQ248257	BQ248257 BQ248257
28	35.4	35.0	694	11	AT108381	AT108381 Zee mays
29	35.4	35.0	831	12	BE889867	BE889867 601512119
30	35.4	35.0	1660	17	AG058676	AG058676 Pan trogl
31	35.2	34.9	535	13	BI597477	BI597477 603243696
32	35.2	34.9	942	13	BC029543	BC029543 Homo sapi
33	35.2	34.9	1092	11	BC029543	BC029543 Homo sapi
34	35.2	34.9	1235	13	BM547617	BM547617 AGENCOURT
35	35.2	34.7	331	13	BQ471625	BQ471625 BQ471625
36	35.2	34.7	545	13	BM488037	BM488037 pgn2n.pk0
37	35.2	34.7	621	9	AJ448803	AJ448803 AJ448803
38	35.2	34.7	1114	17	AG082098	AG082098 Pan trogl
39	34.8	34.5	768	12	BG786329	BG786329 SEMUK006
40	34.6	34.3	686	17	BH576806	BH576806 BGRK62TF
41	34.6	34.3	925	17	CNS0091P	AL053013 Drosophila
42	34.6	34.3	1063	12	BF235933	BF235933 602025614
43	34.4	34.1	612	10	BE418028	BE418028 SC011.C1
44	34.4	34.1	1057	17	CNS04680	AL276477 Tetradon
45	34.2	33.9	368	10	BE398539	BE398539 WHE0004.E

ALIGNMENTS

RESULT 1
AM163063
LOCUS
DEFINITION
IMAGE:2783684.5' similar to SW:ID4_MOUSE P41139 DNA-BINDING PROTEIN
INIBITOR ID-4. [1] ; , mRNA sequence.

ACCESSION
AM163063
VERSION
AM163063.1 GI:6302096

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gaisel, G., Jost, S.,
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Matra, M., Martin,
J., Moore, B., Schellenberg, K., Stepec, M., Tan, F., Theising, B.,
White, X., Wyllie, T., Waterston, R. and Wilson, R.
WashU-MCI human EST Project
Unpublished (1997)
Other ESTs: au91g11.x1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

TITLE
JOURNAL
COMMENT


```

FEATURES
  PROJECT = 'RGP'
  Location/Qualifiers
    1. 512
      /organism="Oryza sativa (Japonica cultivar-group)"
      /cultivar="Nipponbare"
      /db_xref="taxon:33947"
      /clone="E60522"
      /clone_11b="Rice cDNA from immature leaf including apical meristem (under short day condition)"
      /der_stage="Immature leaf including apical meristem (under short day condition)"
BASE COUNT
  78 a 150 c 176 g 107 t 1 others
ORIGIN
  Query Match 36.4%; Score 36.8; DB 9; Length 512;
  Best Local Similarity 61.5%; Pred. No. 55;
  Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 3 GGAGCAGTGTGATCGCGGCGGTGAG;AACGGTGCAGCGCATCGCGGGTTCGC 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 322 GCGGGGGGGGGCGGATGCCATGCGCGCGCTGCGGGGGGCGCATCGCGCGCTGGC 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 63 GCGCGCGCGCGCGCGACCAACGCGGCTGTGATCTTC 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 382 GCGGGGGGTGCTGTGACGCGCGGCTTCACGTGTC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
B0926826 1025 bp mRNA linear EST 20- AUG-2002
LOCUS AGENCOURT_8773285 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6372361
DEFINITION B0926826 B0926826 B0926826.1 GI:22341857
VERSION B0926826
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1025)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strusberg, Ph.D.
Email: cga@bbs.femail.nih.gov
Tissue Procurement: DCTD/DP/PC/Gazdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.lmn.gov
Plate: LITCM2548 row: a column: 02
High quality sequence start: 18
High quality sequence stop: 403.
Location/Qualifiers
  1. 1025
    /organism="Homo sapiens"
    /db_xref="taxon:3606"
    /clone_11b="NIH_MGC_18"
    /clone_11d="NIH_MGC_18"
    /tissue_type="large cell carcinoma"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
    EcoRI; cDNA made by oligo-dT priming. Directionally cloned
    into EcoRI/XhoI sites using the following 5' adaptor:
    GGCGCAGAG(G). Library constructed by Ling Hong in the
    laboratory of Gerald M. Rubin (University of California,
    Berkeley) using 2A'-cDNA synthesis kit (Stratagene) and
    Superscript II RT (Life Technologies). Note: this is a
    NIH_MGC Library."
FEATURES
  SOURCE
    146 a 273 c 498 g 107 t 1 others
BASE COUNT
ORIGIN

```

[illegible]

[illegible]

FEATURES					
Source					
Missouri, Columbia, MO 65211, USA					
Location/Qualifiers					
1..1389					
/organism="Zea mays"					
/db_xref="MaizeDB:638562"					
/db_xref="taxon:4577"					
/clone="PCO144805"					
/clone_1lb="Maize Mapping Project/DuPont Cornsensus Library"					
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACS in conjunction with the Maize Mapping Project"					
BASE COUNT 316 a 383 c 389 g 301 t					
ORIGIN					
Query Match 36.0%; Score 36.4; DB 11; Length 1389;					
Best Local Similarity 61.7%; Pred. No. 64;					
Matches 58; Conservative 0; Mismatches 36; Indels 0; Gaps 0;					
Oy	6	GCAGGTGCTGATCCGGCGGCGTGTGACCAAGCGGTGCGGCGGCGGCGGTTCGCCGC	65		
Db	189	GGAGGAGNAGAAGGAGGAGCGCGCGCGCGCGCGCGCGCATGGCGTGTTGGGATCGGGA	130		
Oy	66	GCGCGCGCTGCGCACCAAGCGGCTGCATGCTCG	99		
Db	129	GGGAGCGCAGCGCAGCAGGCGGTGCGTGTGTCG	96		
RESULT 10					
LOCUS	W29156	1145 bp		mRNA	linear
DEFINITION	mb96a09.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone				
ACCESSION	W29156	IMAGE:337240.5, mRNA sequence.			
VERSION	W29156.1	GI:1309508			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
JOURNAL	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 1145)				
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,F., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.				
	The WashU-HMHI Mouse EST Project				
	Unpublished (1996)				
	Contact: Marra M/Mouse Est Project				
	WashU-HMHI Mouse Est Project				
	Washington University School of Medicine				
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108				
	Tel: 314 286 1800				
	Fax: 314 286 1810				
	Email: mouseest@watsn.wustl.edu				
	This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.				
COMMENT	MGI:218640				
	Seg primer: ETPrimer				
	High quality sequence stop: 338.				
FEATURES					
Source					
1..1145					
/organism="Mus musculus"					
/db_xref="taxon:10090"					
/clone="IMAGE:337240"					
/clone_1lb="Soares mouse p3NMf19.5"					
/dev_stage="19.5 dpc total fetus"					
/_lab_host="DH10B (ampicillin resistant)"					
/_note="Vector: pRT3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Ist strand cDNA					

was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAGTGCAGCGCCGCAATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT 193 a 289 c 388 g 256 t 19 others

Query Match 35.8%; Score 36.2; DB 14; Length 1145;
Best Local Similarity 61.5%; Pred. No. 72;
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 1 CCGAGCAGTGTGATCGCGCGGTGAGACGCGTGCAGCGATCGCGCGGTTTC 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1015 CCGGCG 1074
OY 61 GCGGCG 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1075 GCGGCG 1105

RESULT 11
BE588131 420 bp mRNA linear EST 17-AUG-2000
LOCUS BE588131/c
DEFINITION WHE0657_G04_G042M Secale cereale root tip cDNA library Secale
cereale cDNA clone WHE0657_104_G04, mRNA sequence.
ACCESSION BE588131
VERSION BE588131.1 GI:9841163
KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
1 (bases 1 to 420)
Anderson, O.D., Butler, E., Chao, S., Gustafson, J.P., Han, P.S., Hsieh, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.T., Seaton, C.L. and Tong, J.C.
Title The structure and function of the expressed portion of the wheat
genomes - Root tip cDNA library from rye (Secale cereale)
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: Oanderson@pw.usda.gov

FEATURES
source
1. 420
/organism="Secale cereale"
/db_xref="taxon:4550"
/clone="WHE0657_G04_G04"
/tissue_lib="Secale cereale root tip cDNA library"
/tissue_type="Root tip"
/dev_stage="Seedlings"
/lab_host="E. coli DH12S"
/note="Vector: pSPCOT2; Site_1: SalI; Site_2: NotI; Seeds
were germinated aseptically on filter paper and
transferred to a hydroponic growth system in a growth
chamber when the primary root was 1 cm in length. Plants
were grown in an environmental chamber until harvesting.
The tissue, total RNA, and poly(A) RNA were prepared and
a cDNA library was made (Butler and Gustafson) at
University of Missouri, Columbia. Plasmid DNA

Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: M13 forward primer.
Location/Qualifiers

1. 420
/organism="Secale cereale"
/db_xref="taxon:4550"
/clone="WHE0657_G04_G04"
/tissue_lib="Secale cereale root tip cDNA library"
/tissue_type="Root tip"
/dev_stage="Seedlings"
/lab_host="E. coli DH12S"
/note="Vector: pSPCOT2; Site_1: SalI; Site_2: NotI; Seeds
were germinated aseptically on filter paper and
transferred to a hydroponic growth system in a growth
chamber when the primary root was 1 cm in length. Plants
were grown in an environmental chamber until harvesting.
The tissue, total RNA, and poly(A) RNA were prepared and
a cDNA library was made (Butler and Gustafson) at
University of Missouri, Columbia. Plasmid DNA

preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

BASE COUNT 48 a 221 c 85 g 66 t
ORIGIN

Query Match 35.6%; Score 36; DB 10; Length 420;
Best Local Similarity 70.6%; Pred. No. 84;
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 6 GCAGGTGTGATCGCGCGGTGAGACGCGTGCAGCGATCGCGCGGTTTCGCGC 65
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 GCAGGTGTGATCGCGCGGTGAGACGCGTGCAGCGATCGCGCGGTTTCGCGC 305

OY 66 GCGGCGCG 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 GCGGCGTGC 297

RESULT 12
BJ254408 483 bp mRNA linear EST 08-APR-2002
LOCUS BJ254408
DEFINITION BJ254408 Y. Ogihara unpublished cDNA library, wh_f Triticum
aestivum cDNA clone whf5120 3', mRNA sequence.
ACCESSION BJ254408
VERSION BJ254408.1 GI:20078766
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
1 (bases 1 to 483)
Ogihara, Y. and Murai, K.
Title Expressed genes in Triticum aestivum
Unpublished (2002)
Contact: Tadashi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES
source
1. 483
/organism="Triticum aestivum"
/db_xref="taxon:4565"
/clone="whf5120"
/tissue_lib="Y. Ogihara unpublished cDNA library, wh_f"
/tissue_type="Spike at flowering date"
/dev_stage="Seedlings"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown under
hydroponic conditions at UC Davis, salt stressed for 12
hours, and for 7 days, then dissected and frozen (Akhunov
in J Dvorak Lab). Total RNA was prepared from shoot
tissue, equal quantities of RNA were pooled from the two
samples, polyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo
excised to give phagescript phagemids in the T3 Close lab
at the University of California, Riverside (Akhunov, Chin
Choi, Close, Fenton, Kianian, Otto, Simons, Zhang).
Plasmid DNA preparations and DNA sequencing were
performed in the OD Anderson lab (all other authors)."

BASE COUNT 91 a 161 c 129 g 98 t 4 others

Query Match 35.6%; Score 36; DB 13; Length 483;
Best Local Similarity 63.5%; Pred. No. 83;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

OY 15 GATCGCGCGGTGAGACGCGTGCAGCGATCGCGCGGTTTCGCGCGCGCGCT 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db	319	GAGCTCCGCGGTGTAGCGCGCGCTGCACATGTAAGCTGCACAGGTACGGGCCCGCAGGGGT	378
OY	75	GGCAGCACAAGCGCGCTGCATCTCTCG	99
Db	379	GGCATCAGCGCGCTCTCTCTCTCG	403
RESULT 13			
LOCUS	A0427466		
DEFINITION	A0427466	554 bp	DNA
ACCESSION	A0427466		linear
VERSION	A0427466.1		GS3 24-MAR-1999
KEYWORDS	GS3.		genomic clone 2555p11,
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 554)		
JOURNAL	Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and		
COMMENT	Venter,J.C.		
	Use of BAC End Sequences from Caltech Libraries for Sequence-Ready		
	Map Building		
	Unpublished (1997)		
	Other_GSSs: CITBI-E1-2555p11.TR		
	Contact: Shaying Zhao, William Nierman, Mark Adams		
	Department of Eukaryotic Genomics		
	The Institute for Genomic Research		
	7912 Medical Center Dr., Rockville, MD 20850		
	Tel: 301 838 0200		
	Fax: 301 838 0208		
	Email: hoe@tigr.org		
	Clones are available from Research Genetics (info@resgen.com). BAC		
	end search page:		
	http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.		
	Seq primer: M13-21		
	Class: BAC ends.		
FEATURES	Location/Qualifiers		
source	1..554		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="2555p11"		
	/clone_1lb="CITBI-E1"		
	/sex="male"		
	/cell_type="sperm"		
	/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;		
BASE COUNT	91 a 176 c 221 g 66 t		
ORIGIN			
Query Match	35.2%; Score 35.6; DB 17; Length 554;		
Best Local Similarity	64.6%; Pred. No. 1e+02; Mismatches 29; Indels 0; Gaps 0;		
Matches	53; Conservative 0;		
OY	6	GCAGGTGTGATCGCGGCGGTGACCAAGCGGTGCAGCGATCGCGCGGCTTCGCGGC	65
Db	390	GCAGAGGTGGGCGCGCGCGGACCCGCGCGGCGCGCGCGCGCGGAGTACGCGGT	449
OY	66	GCAGGTGTGATCGCGGCGGTGACCAAGCGG	87
Db	450	CCGGGACTTACCGGGAACCG	471
RESULT 14			
LOCUS	BM656407		
DEFINITION	BM656407	596 bp	mRNA
ACCESSION	BM656407		linear
VERSION	BM656407.1		EST 26-FEB-2002
KEYWORDS	EST.		
SOURCE	African malaria mosquito.		

[illegible]

Plate: NU01004AAO row: G column: 10
Seq primer: M13 Reverse.
Location/Qualifiers
1. 611
source

FEATURES
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7163"
/clone_11b="A.Gam.ad.cDNA.blood1"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mt4.org)"
BASE COUNT 127 a 180 c 180 g 124 t
ORIGIN

Query Match 35.2%; Score 35.6; DB 13; Length 611;
Best Local Similarity 60.2%; Pred. No. 1e+02;
Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 4 GAGCAGGTGTGATCGCGGCGGTGAGCAACGCTGCAGCGCATCGCGGCGGTTCGCG 63
|||||
DB 465 GAGCGGGTGTCTCCCTCCATCAGCAGCGAAGTGTGAAGGCGGTGCGCCCATTCGAT 524
|||||
QY 64 GCGCGCGCGCTGCGCACCACCAAGCGGCTGCATGTCGCA 101
|||||
DB 525 GCGGGCGAGCTGATCAGCAGCGAGATGATGTCGCA 562
|||||

Search completed: November 5, 2002, 16:17:34
Job time: 1119.44 secs

APPLICANT: Sleeman, Matthew
APPLICANT: Abernethy, Nevin
APPLICANT: Onrust, Rene
APPLICANT: Kumble, Anand
APPLICANT: Murlison, Greg
TITLE OF INVENTION: Compositions Isolated From Stromal Cells
FILE REFERENCE: 11000.1037c3
CURRENT APPLICATION NUMBER: US/09/823,038A
CURRENT FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 45
LENGTH: 925
TYPE: DNA
ORGANISM: Mouse
US-09-823-038A-45

Query Match 32.1%; Score 32.4; DB 10; Length 925;
Best Local Similarity 60.0%; Pred. No. 3.1;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 2 CGGAGCAGGTGTCGATCGGGCGGTGAGCAGCGGTGCAGCGCGGCGGTTG 61
DB 395 CGGCCAGGCGCTGTGTGCGCGGTGTGCGCGCGCTGTGTGCGCGCGG 454
OY 62 CGGCGCGCGCGCTGTGTGCGCGCGCGCGCGCTGTGTGCGCGCGG 91
DB 455 TGTCTGAGCTTCTGTGTGCGCTGTGTGCGCGCGCGCGCGCGCGCG 484

RESULT 3

US-09-917-800A-1719/c
Sequence 1719, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917,800A
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1719
LENGTH: 1408
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_024125
US-09-917-800A-1719

Query Match 32.1%; Score 32.4; DB 10; Length 1408;
Best Local Similarity 60.0%; Pred. No. 3;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 2 CGGAGCAGGTGTCGATCGGGCGGTGAGCAGCGGTGCAGCGCGGCGGTTG 61
DB 650 CGAAGCAGCGCGCGCGCGCGCGCTGTGTGCGCGCGCGCGCGCGCGG 591
OY 62 CGGCG 91
DB 590 ACGACGACGACGTGTGACAGCGCTGCGCGCTGC 561

RESULT 4

US-09-789-836-10/c
Sequence 10, Application US/09789836
Patent No. US20020082204A1
GENERAL INFORMATION:
APPLICANT: BRIGHAM, KENNETH L.
APPLICANT: STECENKO, ARLENE A.
APPLICANT: SEALY, LINDA
TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
FILE REFERENCE: N-6977
CURRENT APPLICATION NUMBER: US/09/789,836
CURRENT FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/183,584
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 1739
TYPE: DNA
ORGANISM: Rattus sp.
US-09-789-836-10

Query Match 32.1%; Score 32.4; DB 10; Length 1739;
Best Local Similarity 60.0%; Pred. No. 3;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 2 CGGAGCAGGTGTCGATCGGGCGGTGAGCAGCGGTGCAGCGCGGCGGTTG 61
DB 736 CGAAGCAGCGCGCGCGCGCGCGCTGTGTGCGCGCGCGCGCGCGCGG 677
OY 62 CGGCG 91
DB 676 ACGACGACGACGTGTGACAGCGCTGCGCGCTGC 647

RESULT 5

US-09-867-550-233
Sequence 233, Application US/09867550
Patent No. US20020082206A1
GENERAL INFORMATION:
APPLICANT: Leach, Martin D.
APPLICANT: Mehraban, Foad,
APPLICANT: Conley, Pamela
APPLICANT: Law, Debbie
APPLICANT: Topper, James
TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells a
FILE REFERENCE: 21402-013 (Cura-313)
CURRENT APPLICATION NUMBER: US/09/867,550
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: USSN 60/208,427
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 2125
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 233
LENGTH: 379
TYPE: DNA
ORGANISM: Homo sapiens
US-09-867-550-233

QY 3 GGACACAGTGGTATCCCGGCGCTGGAG:CAAGGGGTGACAGCGAATGGCGCGGGTTCG 62
 1019 GGAACGCGTGGCCCGCGACCGCTGGA:GAGCTCGGGCTGCGGGAATGCCCGGTCTCG 1078
 QY 63 GGGCGCGCGCTGCGGACCAAGCGCTCAATGTC 96
 Db 1079 GGTGCGCGGCGAGACACCAACCCCGTCTGGGTC 1112

RESULT 10
US-09-726-397A-3/c
; Sequence 3, Application US/09726397A
Date: 09/09/2009 10:11:11

```

1 patient NO. US2002014201A1
2 GENERAL INFORMATION:
3 APPLICANT: PARISH, TANYA
4 APPLICANT: SMITH, DEBBIE
5 TITLE OF INVENTION: VACCINE
6 FILE REFERENCE: 117-329
7 CURRENT APPLICATION NUMBER: US/09/726,397A1
8 CURRENT FILING DATE: 2000-12-01
9 PRIOR APPLICATION NUMBER: 0014845.20
10 PRIOR FILING DATE: 2000-06-16
11 NUMBER OF SEQ ID NOS: 4
12 SOFTWARE: PatentIn Ver. 2.1
13 SEQ ID NO 3
14 LENGTH: 2366
15 TYPE: DNA
16 ORGANISM: Mycobacterium tuberculosis
17 FEATURE:
18 NAME/KEY: misc_feature
19 LOCATION: (1838)..(2090)
20 OTHER INFORMATION: unknown
21 NAME/KEY: misc_feature
22 LOCATION: (2103)..(2108)
23 OTHER INFORMATION: unknown
24
25 US-09-726-397A-3

```

Query Match	31.3%	Score	31.6	DB	10	Length	2366
Best Local Similarity	58.5%	Pred. No.	4.6				
Matches	55	Conservative	0	Mismatches	39	Indels	0
						Gaps	0

[illegible]

```

RESULT 11
US-09-894-998-35/C
Sequence 35, Application US/0984998
Patent No. US200200610A1
GENERAL INFORMATION:
APPLICANT: Hosken, Nancy Ann
APPLICANT: Craig H. Day
APPLICANT: David C. Dillon
APPLICANT: McGowan, Patrick
APPLICANT: Steach, Paul R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
FILE REFERENCE: 210121.538
CURRENT APPLICATION NUMBER: US/09/894,998
CURRENT FILING DATE: 2001-06-28
NUMBER OF SEQ. ID NOS: 64
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ. ID NO 35
LENGTH: 2481
TYPE: DNA
ORGANISM: HSV-2
US-09-894-998-35

```

Query Match	30.7%	Score 31;	DB 10;	Length 2481;
Best Local Similarity	64.8%;	Pred. No. 6.4;		
Matches 46;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0

QY	3	GGACGAGTGTGATGTC	CGGCGCTGGAGCAAGCGGTG	CAAGCGCGCGCGGTTGCG	62
Db	1815	GGCGGAGAGGAGAGAGCGG	AGGAGGAGAGCGCGGAGGAG	AGCGCGCGCGGCGGCGGCG	1756
QY	63	GGCGCGCGCGCG	73		
Db	1755	GGCGTGGGACG	1745		

Oy	63	GGCGCGCGCG	73
Db	1755	GGCCTGGGACG	1745

RESULT 12
US-09-864-761-22031
; Sequence 22031, Application US/09864761

Patent NO. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Weosheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: A6001ca-X-1

```

CURRENT APPLICATION NUMBER: US/05/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 22031
LENGTH: 372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006385.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5,
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.9

```

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822,830A
CURRENT FILING DATE: 2001-03-29

; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 19
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-19

Query Match 30.3%; Score 30.6; DB 10; Length 1362;
Best Local Similarity 60.0%; Pred. No. 8.2;
Matches 51; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 2 CGGACGAGTGTGATCGCGCGGTGGAGCAACGCGTGCAGCGCATCGCGCGGGTTTCG 61
DB 607 CGGCGCGCGTGTGTGCGCAGCGGTGGCGGTGCGCAGTGTGTGGCGGTGGCGGTGGCAG 548
QY 62 CGGCGCGCGCGGTGCGCAGCAGCG 86
DB 547 CAGCGGCGCACAGCCCCCAGAG 523

Search completed: November 5, 2002, 22:59:16
Job time : 89.2537 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_18300_18400

Perfect score: 101

Sequence: 1 ccgagcagcgtgtgtatcgc.....aagcgcgtcatgtctgcga 101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_NA.*
2: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	71989	4 US-09-443-501A-2	Sequence 2, Appli
2	99.4	98.4	68750	3 US-09-335-409-1	Sequence 1, Appli
3	99.4	98.4	68750	4 US-09-568-102-1	Sequence 1, Appli
4	99.4	98.4	68750	4 US-09-567-969-1	Sequence 1, Appli
5	99.4	98.4	68750	4 US-09-568-480-1	Sequence 1, Appli
6	99.4	98.4	68750	4 US-09-568-486-1	Sequence 1, Appli
7	99.4	98.4	68750	4 US-09-568-472-1	Sequence 1, Appli
8	99.4	98.4	68750	4 US-09-567-899-1	Sequence 1, Appli
9	51.4	50.9	15872	4 US-09-105-537-1	Sequence 1, Appli
10	51.4	50.9	43280	2 US-08-804-227C-1	Sequence 1, Appli
11	43.4	43.0	33529	4 US-09-144-085-3	Sequence 3, Appli
12	38.2	37.8	2721	6 5215861-2	Patent No. 5215861
13	38.2	37.8	8438	1 US-07-945-283-1	Sequence 1, Appli
14	36.8	36.4	4403765	4 US-09-103-840A-2	Sequence 1, Appli
15	36.8	36.4	4411529	4 US-09-103-840A-1	Sequence 2, Appli
16	36.2	35.8	1505	1 US-07-915-246-1	Sequence 1, Appli
17	35.4	35.0	80161	3 US-09-036-987A-1	Sequence 1, Appli
18	35.4	35.0	80161	3 US-09-370-700-1	Sequence 1, Appli
19	35.2	34.9	1088	3 US-09-040-285A-1	Sequence 1, Appli
20	34.8	34.5	15393	3 US-09-453-702B-191	Sequence 191, App
21	34.2	33.9	44377	2 US-08-804-227C-7	Sequence 7, Appli
22	34.2	33.9	44377	2 US-08-804-198-1	Sequence 1, Appli
23	33	32.7	1030	3 US-08-858-003-2	Sequence 2, Appli
24	33	32.7	1030	3 US-09-078-166-2	Sequence 2, Appli
25	33	32.7	1030	3 US-08-997-467-2	Sequence 2, Appli
26	32.8	32.5	7812	4 US-09-368-590-1	Sequence 1, Appli
27	31.8	31.5	3205	4 US-09-046-894-34	Sequence 34, Appli

28	31.4	31.1	20235	1 US-07-642-734C-3	Sequence 3, Appli
29	31.4	31.1	20235	3 US-08-439-009A-3	Sequence 3, Appli
c 30	31.2	30.9	4411529	4 US-09-103-840A-1	Sequence 1, Appli
31	31	30.7	1144	1 US-08-014-943A-1	Sequence 1, Appli
32	31	30.7	1144	1 US-08-486-421-2	Sequence 2, Appli
33	31	30.7	1144	1 US-08-470-911-2	Sequence 2, Appli
34	31	30.7	1144	2 US-08-486-809-2	Sequence 2, Appli
c 35	31	30.7	1336	4 US-09-180-109A-28	Sequence 28, Appli
c 36	31	30.7	1420	4 US-09-180-109A-23	Sequence 23, Appli
37	31	30.7	2214	3 US-08-864-038A-1	Sequence 1, Appli
38	31	30.7	3331	3 US-08-864-038A-2	Sequence 1, Appli
39	31	30.7	3331	3 US-08-864-038A-4	Sequence 4, Appli
c 40	30.8	30.5	1500	4 US-09-593-711A-10	Sequence 10, Appli
41	30.8	30.5	1748	3 US-09-100-730-1	Sequence 1, Appli
42	30.8	30.5	1926	4 US-09-249-585A-2	Sequence 2, Appli
43	30.8	30.5	2580	3 US-09-050-863-2	Sequence 2, Appli
44	30.8	30.5	2580	4 US-09-359-081-2	Sequence 2, Appli
c 45	30.8	30.5	5452	2 US-09-130-114-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-443-501A-2
: Sequence 2, Application US/09443501A
: Patent No. 6303342
: GENERAL INFORMATION:
: APPLICANT: Kusan Biosciences, Inc.
: APPLICANT: Julien, Bryan
: APPLICANT: Katz, Leonard
: APPLICANT: Khosla, Chaitan
: APPLICANT: Tang, Li
: APPLICANT: Ziemann, Rainer
: TITLE OF INVENTION: Recombinant Methods and Materials for Producing
: TITLE OF INVENTION: Epothilone and Epothilone Derivatives
: FILE REFERENCE: 30062-20031.00
: CURRENT APPLICATION NUMBER: US/09/443,501A
: PRIOR FILING DATE: 1999-11-19
: PRIOR APPLICATION NUMBER: US 60/130,560
: PRIOR FILING DATE: 1999-04-22
: PRIOR APPLICATION NUMBER: US 60/122,620
: PRIOR FILING DATE: 1999-03-03
: PRIOR APPLICATION NUMBER: US 60/119,386
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: US 60/109,401
: PRIOR FILING DATE: 1998-11-20
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 71989
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match      100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.2e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CCGGAGCAGGTGTCATATCGCGGCGCTGAGACAAAGCGTGCAGCGCGGCGGTTC 60
Db      18300 CCGGAGCAGGTGTCATATCGCGGCGCTGAGACAAAGCGTGCAGCGCGGCGGTTC 18359

QY      61 GCGGCGCGCGCGGTGCGACCAAGCGGTGCATGTCTCGCA 101
Db      18360 GCGGCGCGCGCGGTGCGACCAAGCGGTGCATGTCTCGCA 18400

RESULT 2
US-09-335-409-1
: Sequence 1, Application US/09335409
```

```
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match          98.4%; Score 99.4; DB 3; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 60
    |||||||
DB 23912 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 23971
```

```
QY 61 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 101
    |||||||
DB 23972 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 24012
```

```
RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-102-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 60
    |||||||
DB 23912 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 23971
```

```
QY 61 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 101
    |||||||
DB 23972 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 24012
```

```
RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-567-969-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 60
    |||||||
DB 23912 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 23971
```

```
QY 61 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 101
    |||||||
DB 23972 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 24012
```

```
RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1
```

```
Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 60
    |||||||
DB 23912 CCGGAGCAGGTGTGATCGCGGCGGTGAGCAAGCGGTGCAGCGCATCGCGGGGTTTC 23971
```

```
QY 61 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 101
    |||||||
DB 23972 GCGGCGCGCGGCGGTGCACCAAGCGGCTGATGTTCTCGCA 24012
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAGACAGGTGTGATCCGGCGCGGTGAGCAAGCGTGACGCGGCGGCGGTTTC 60
DB 23912 CCGAGACAGGTGTGATCCGGCGCGGTGAGCAAGCGTGACGCGGCGGCGGTTTC 23971
QY 61 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 101
DB 23972 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 24012

RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAGACAGGTGTGATCCGGCGCGGTGAGCAAGCGTGACGCGGCGGCGGTTTC 60
DB 23912 CCGAGACAGGTGTGATCCGGCGCGGTGAGCAAGCGTGACGCGGCGGCGGTTTC 23971
QY 61 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 101
DB 23972 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 24012
```

```
DB 23972 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 24012

RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match          98.4%; Score 99.4; DB 4; Length 68750;
Best Local Similarity 99.0%; Pred. No. 5e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAGACAGGTGTGATCCGGCGCGGTGAGCAAGCGTGACGCGGCGGCGGTTTC 60
DB 23912 CCGAGACAGGTGTGATCCGGCGCGGTGAGCAAGCGTGACGCGGCGGCGGTTTC 23971
QY 61 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 101
DB 23972 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 24012

RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match          50.9%; Score 51.4; DB 4; Length 15872;
Best Local Similarity 72.0%; Pred. No. 0.00037;
Matches 67; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 GGTGGTATCGCGGCGGTGAGCAAGCGGATCGCGGCGGCGGCGGCGGCGGCGG 68
DB 10606 GGTGGTATCTCCGCGCGGAGAGCGGTGCGGAGCGCGCGCGCGCGGAGC 10665
QY 69 CGCGTGCGCACCAAGCGGTCGATGTCTCGCA 101
DB 23972 GCGGCGCGCGCGCGCGCGCACCAAGCGGTCGATGTCTCGCA 24012
```

Db 10666 AGCGCGCCGACACAGCGCGCTCCGCGTCGCCA 10698

RESULT 10

US-08-804-227C-1
Sequence 1, Application US/08804227C
Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstosa, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 43280 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 816..14234

FEATURE:

NAME/KEY: CDS

LOCATION: 14351..19945

FEATURE:

NAME/KEY: CDS

LOCATION: 20010..31199

FEATURE:

NAME/KEY: CDS

LOCATION: 31232..36067

FEATURE:

NAME/KEY: CDS

LOCATION: 36249..41774

US-08-804-227C-1

Query Match

Best Local Similarity 50.98; Score 51.4; DB 2; Length 43280;
Matches 67; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db 9 GGTGGTATCGCGGCGTGGAGAGCGTCAGCGCATCCGGCGGCGTTCGGCGCGC 68

Db 38468 GGTGGTATCGCGGCGTGGAGAGCGTCAGCGCATCCGGCGGCGTTCGGCGCGC 38527

Qy 69 CGCGGTGCGACACAGCGGTGATGTCATGCA 101

Db 38528 AGCGCGCGCGACACAGCGGTGATGTCATGCA 38560

RESULT 11

US-09-144-085-3
Sequence 3, Application US/09144085
Patent No. 6280999

GENERAL INFORMATION:

APPLICANT: Gustafsson, Claes
APPLICANT: Betsch, Mary C.
APPLICANT: Ashley, Gary
APPLICANT: Julien, Bryan

TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
FILE REFERENCE: 30062-20020.20

CURRENT APPLICATION NUMBER: US/09/144, 085
EARLIER FILING DATE: 1998-08-31
EARLIER APPLICATION NUMBER: 09/010, 809

NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 3

LENGTH: 33529
TYPE: DNA
ORGANISM: Sorangium cellulosum

US-09-144-085-3

Query Match

Best Local Similarity 43.0%; Score 43.4; DB 4; Length 33529;
Matches 65; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1 CCGAGCAGGTGTGATCGCGGCGGTGAGACACCGGTGACAGCGGATCGCGGCGGTTG 60

Db 22758 CCGATGCGAGAGCGGTGTGAGAGCGGCGGCGGCGGCGGTGCGGCGGCGTGTG 22817

Qy 61 CGCGCGCGCGCGGTGCGCGACACCGCGGTGATGTCGCA 101

Db 22818 GAGCGCGAGGCGCGCGCGCGCGCGGTGCTGTGTCGCA 22858

RESULT 12

5215881-2/c

APPLICANT: CHEUNG, ANDREW K.

TITLE OF INVENTION: PSEUDORABIES DIAGNOSIS PROBES

NUMBER OF SEQUENCES: 3

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/537, 855

FILING DATE: 13-JUN-1990

SEQ ID NO:2

LENGTH: 2721

5215881-2

Query Match

Best Local Similarity 37.8%; Score 38.2; DB 6; Length 2721;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 2 CGGAGCAGGTGTGATCGCGGCGGTGAGACACCGGTGACAGCGGATCGCGGCGGTTG 61

Db 787 CGCGCGCGCGCGAGTGTGCGCGGTGTGAGAGCGGCGGCGCGCGCGCGGCGGCGT 728

Qy 62 CGCGCGCGCGCGTGTGCGACACAGCGGTGATGTCGC 100

Db 727 CGCGCGCGCGAGATGTCCCGGTCTCTCTCTCCGC 689

RESULT 13

US-07-945-283-1/c
Sequence 1, Application US/07945283
Patent No. 5352596

GENERAL INFORMATION:

APPLICANT: Cheung, Andrew K.
APPLICANT: Wesley, Ronald D.

TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants


```

; TITLE OF INVENTION: Involving The EP0 and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 535256th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8438 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: pseudorabies virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 622..6495
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1099, "g")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1267, "t")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1381, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1566, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(7010, "g")
; US-07-945-283-1

Query Match 37.8%; Score 38.2; DB 1; Length 8438;
Best Local Similarity 61.6%; Pred. No. 0.37;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 2 CGGAGCAGGTGCTGATCCGGCGCTGAGCAAGCGGTGCGGCGGGGCTTCG 61
DB 2130 CGGCGCGCGCGAGGTGGCGGGGTGTGGAAGCGGCGGCGGCGGCGGAGAGGCT 2071
QY 62 CGGCGCGCGCGCTGCGCACCAAGCGGCTGCATGCTTCGC 100
DB 2070 CGGCGCGCGAGGATGCTCCCGCTTCTCTCCCGC 2032

RESULT 14
US-09-103-840A-2
; Sequence 2, Application us/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
; US-09-103-840A-2

Query Match 36.4%; Score 36.8; DB 4; Length 4403765;
Best Local Similarity 63.6%; Pred. No. 0.56;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2 CGGAGCAGGTGCTGATCCGGCGCTGAGCAAGCGGTGCGGCGGGGCTTCG 61
DB 835263 CGACGGCGGTGCTGGCGCGCGCGGATGCTGCGTGCATCGCGCGCTTTTG 835322
QY 62 CGGCGCGCGCGCTGCGCACCAAGCGGCT 89
DB 835323 GGGCGCAGCGCCAGGCTTATCAGGCGCT 835350

RESULT 15
US-09-103-840A-1
; Sequence 1, Application us/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
; US-09-103-840A-1

Query Match 36.4%; Score 36.8; DB 4; Length 4411529;
Best Local Similarity 63.6%; Pred. No. 0.56;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2 CGGAGCAGGTGCTGATCCGGCGCTGAGCAAGCGGTGCGGCGGGGCTTCG 61
DB 833085 CGACGGCGGTGCTGGCGCGCGCGGATGCTGCGTGCATCGCGCGCTTTTG 833144
QY 62 CGGCGCGCGCGCTGCGCACCAAGCGGCT 89
DB 833145 GGGCGCAGCGCCAGGCTTATCAGGCGCT 833172

Search completed: November 5, 2002, 13:38:04
Job time : 2032.74 secs
```



```
FT      /tag- d
FT      /note- "encodes ket::de synthase (KS-Y) of the loading
FT      domain"
FT      3621..4661
FT      /tag- e
FT      /note- "encodes acyl. transferase (AT) of the loading
FT      domain"
FT      4917..5810
FT      /tag- f
FT      /note- "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      5856..6155
FT      /tag- g
FT      /note- "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      6260..10493
FT      /tag- h
FT      /label- epob_gene
FT      /note- "encodes module 1, the NRPS module"
FT      2031..3548
FT      /tag- i
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      2031..3548
FT      /tag- j
FT      /note- "encodes condensation domain C2 of the NRPS
FT      module"
FT      6861..6887
FT      /tag- k
FT      /note- "encodes heterocyclization signature sequence"
FT      6861..6887
FT      /tag- l
FT      /note- "encodes condensation domain C4 of the NRPS
FT      module"
FT      7358..7366
FT      /tag- m
FT      /note- "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      7898..7921
FT      /tag- n
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      7898..7921
FT      /tag- o
FT      /note- "encodes adenylation domain A1 of the NRPS module"
FT      8261..8308
FT      /tag- p
FT      /note- "encodes adenylation domain A3 of the NRPS module"
FT      8411..8422
FT      /tag- q
FT      /note- "encodes adenylation domain A4 of the NRPS module"
FT      8861..8905
FT      /tag- r
FT      /note- "encodes adenylation domain A6 of the NRPS module"
FT      8966..8983
FT      /tag- s
FT      /note- "encodes adenylation domain A7 of the NRPS module"
FT      9090..9179
FT      /tag- t
FT      /note- "encodes adenylation domain A8 of the NRPS module"
FT      9183..9992
FT      /tag- u
FT      /note- "encodes oxidation region for forming thiazole"
FT      10121..10138
FT      /tag- v
FT      /note- "encodes adenylation domain A10 of the NRPS
FT      module"
FT      10261..10306
FT      /tag- w
FT      /note- "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      10639..16137
FT      /tag- x

FT      /label- epoc_gene
FT      /note- "encodes module 2"
FT      10654..12033
FT      /tag- y
FT      /note- "encodes KS2, the KS domain of module 2"
FT      12250..13287
FT      /tag- z
FT      /note- "encodes AT2, the AT domain of module 2"
FT      13327..13899
FT      /tag- aa
FT      /note- "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      14962..15736
FT      /tag- ab
FT      /note- "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      15763..16008
FT      /tag- ac
FT      /note- "encodes ACP2, the ACP domain of module 2"
FT      16134..37907
FT      /tag- ad
FT      /label- epod_gene
FT      /note- "encodes modules 3-6"
FT      16425..17606
FT      /tag- ae
FT      /note- "encodes KS3"
FT      17817..18857
FT      /tag- af
FT      /note- "encodes AT3"
FT      19581..20396
FT      /tag- ag
FT      /note- "encodes KR3"
FT      20424..20642
FT      /tag- ah
FT      /note- "encodes ACP3"
FT      20706..22082
FT      /tag- ai
FT      /note- "encodes KS4"
FT      22296..23336
FT      /tag- aj
FT      /note- "encodes AT4"
FT      24069..24647
FT      /tag- ak
FT      /note- "encodes KR4"
FT      24867..25151
FT      /tag- al
FT      /note- "encodes ACP4"
FT      25203..26576
FT      /tag- am
FT      /note- "encodes KS5"
FT      26793..27883
FT      /tag- an
FT      /note- "encodes AT5"
FT      27966..28574
FT      /tag- ao
FT      /note- "encodes DHS"
FT      29433..30287
FT      /tag- ap
FT      /note- "encodes ER5"
FT      30321..30869
FT      /tag- aq
FT      /note- "encodes KR5"
FT      31077..31373
FT      /tag- ar
FT      /note- "encodes ACP5"
FT      31440..32807
FT      /tag- as
FT      /note- "encodes KS6"
FT      33018..34067
FT      /tag- at
FT      /note- "encodes AT6"
FT      34107..34676
FT      /tag- au
```

```

FT      /note= "encodes DH6"
FT      35760..36641
FT      /tag= av
FT      /note= "encodes ER6"
FT      36705..37256
FT      /tag= aw
FT      /note= "encodes KR6"
FT      37470..37769
FT      /tag= ax
FT      /note= "encodes ACP6"
FT      37912..49308
FT      /tag= ay
FT      /label= epoe.gene
FT      /note= "encodes modules 7 and 8"
FT      38014..39375
FT      /tag= az
FT      /note= "encodes KS7"
FT      39589..40626
FT      /tag= ba
FT      /note= "encodes AT7"
FT      41341..41922
FT      /tag= bb
FT      /note= "encodes KR7"
FT      42181..42423
FT      misc_RNA

Query Match      100.0%; Score 101; DB 21; Length 71989;
Best local similarity 100.0%; Pred. NO. 4,1e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 CCGGAGCAGTGTGATCGCGGGCGTGGAGCAAGCGGTGCGAGCGCGGGGCTTC 60
        |||
Db      18300 CCGGAGCAGTGTGATCGCGGGCGTGGAGCAAGCGGTGCGAGCGCGGGGCTTC 18359
Oy      61 GCGGCGCGCGGCGTGCACCAAGCGGCTGCATGTCTGCCA 101
        |||
Db      18360 GCGGCGCGCGGCGTGCACCAAGCGGCTGCATGTCTGCCA 18400

RESULT 2
AA255887
ID      AA255887 standard; DNA; 68750 BP.
XX      AC      AA255887;
XX      DT      10-APR-2000 (first entry)
XX      DE      Sorangium cellulosum 68.75 kb contig.
XX      KW      Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
XX      KW      anticancer; ds.
XX      OS      Sorangium cellulosum.
XX      FH      Location/Qualifiers
FT      CDS      1..1826
                /tag= a
                /partial
                /product= "Partial Orf 1 protein (AAV58580)"
                /note= "No initiation codon given in the specification"
FT      CDS      complement (1900..3171)
                /tag= b
                /product= "Orf 2 protein (AAV58581)"
                3415..5556
                /tag= c
                /product= "Orf 3 protein (AAV58582)"
                complement (5612..5992)
                /tag= d
                /product= "Orf 4 protein (AAV58583)"
                6226..6675
                /tag= e
                /product= "Orf 5 protein (AAV58584)"
                7610..11875
                /tag= f
FT      CDS

```

```

FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /partial
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
FT      MO9966028-A2.
XX      PD      23-DEC-1999.
XX      PF      16-JUN-1999; 99WO-EP04171.
XX      PR      18-JUN-1998; 98US-0099504.
XX      PR      24-SEP-1998; 98US-0101631.
XX      PR      05-FEB-1999; 99US-0118906.
XX      PA      (NOVS ) NOVARTIS AG.
XX      PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX      PI      Schnupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;

```

```
XX WPI; 2000-097711/08.
DR AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR P-PSDB; AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR AAY58592, AAY58593, AAY58594.

PT New isolated epothilone synthase genes, used for the recombinant
PT production of epothilone for use in cancer therapy -
XX
XX Claim 14; Page 87-104; 174pp; English.

CC This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC comprising 22 open reading frames (ORFs) and includes genes encoding
CC proteins involved in the biosynthesis of epothilones. Epothilones A and
CC B are 16-membered macrocyclic polyketides with an acylglycine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AAY58575-58578) are involved in polyketide backbone
CC formation. EPO F (AAY58579) is an epothilone macroactone oxidase, and
CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epothilones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epothilones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epothilones as anticancer agents, they are problematical to produce on a
CC large scale. Epothilones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epothilones. The nucleic acids of the invention may be
CC used for the recombinant production of epothilones in a heterologous host
CC that is more amenable to fermentation.
XX
XX Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
SQ

Query Match 98.4%; Score 39.4; DB 21; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9.6e-15;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CCGGAGCAGGTGATCCGGGCGGTGAGCAAGCGGTGAGCGGCGGGGTTTC 60
DB 23912 CCGGAGCAGGTGATCCGGGCGGTGAGCAAGCGGTGAGCGGCGGGGTTTC 23971
OY 61 GCGGCGCGCGGCTGCGGCAACAAGCGGCTGCAATCTCTCGCA 101
DB 23972 GCGGCGCGCGGCGCGGCAACAAGCGGCTGCAATCTCTCGCA 24012

RESULT 3
AAT68715
ID AAT68715 standard; DNA: 15872 BP.
XX
AC AAT68715;
XX
DT 01-SEP-1997 (first entry)
XX
DE Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW polyhydroxybutyrate; biodegradable polymer; vep gene;
KW metabolic engineering; ss.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 20..13912
FT /tag= b
FT CDS 14056..14136
```

```
FT /tag= b
FT CDS 14148..15827
FT /tag= c
XX
XX MO9722711-AL.
XX
XX 26-JUN-1997.
XX
XX 18-DEC-1996; 96WO-US20119.
XX
XX 19-DEC-1995; 95US-0008847.
XX
XX (MIND ) UNIV MINNESOTA.
XX
XX Sherman DH, Williams MD, Xue Y;
XX
XX WPI: 1997-341701/31.
XX
XX P-PSDB; AAW19629-30 AND AAW00918.
XX
XX Expression cassettes for production of polyhydroxyalkanoate(s) -
XX provide wide range of biodegradable polymers for medical or
XX industrial use
XX
XX Claim 54; Fig 23; 91pp; English.
XX
XX Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
XX synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
XX also AAW19629-30 and AAW00918). It contains 5 PKS modules, with a 5'
XX loading module and a 3' end domain. Each of the sequenced modules
XX includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
XX reductase and an acyl carrier protein domain. The gene cluster was
XX cloned using a heterologous hybridisation strategy from a genomic
XX DNA library. A novel expression cassette encoding the first module
XX from the vep gene cluster and module 7 from the Streptomyces tyIP
XX gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
XX activity and can be used for PHA produ. in host (esp. insect) cells
XX for use as a biodegradable polymer.
XX
XX Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;
SQ

Query Match 50.9%; Score 51.4; DB 18; Length 15872;
Best Local Similarity 72.0%; Pred. No. 0.0014;
Matches 67; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

OY 9 GGTGTGATCCGCGGCGGTGAGCAAGCGGTGAGCGGCGGGGTTTCGCGCGG 68
DB 10606 GGTGTGATCTCCGGCGGAGAGACCGGTGCGGCGGCGGCGGCGGCGGCGGAGC 10665
OY 69 CCGCGTCCGACCAACCGGCTGCATCTCTCGCA 101
DB 10666 AGCGCCGCGCACACGAGCGGCTCGCGTCGCGCA 10698

RESULT 4
AAZ87283
ID AAZ87283 standard; DNA: 15872 BP.
XX
AC AAZ87283;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW desosamine biosynthesis; macroide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
FH Key Location/Qualifiers
FT CDS 20..13912
```

FT	/ftag= a
FT	/product= "vcp ORF 1 amino acid sequence #1 (AAV77177) "
FT	CDS
FT	14056..14151
FT	/tag= b
FT	/product= "vcp ORF 1 amino acid sequence #3 (AAV77199) "
FT	CDS
FT	14167..15827
FT	/tag= c
FT	/product= "vcp ORF 1 amino acid sequence #2 (AAV77178) "
XX	
PN	WC200000620-A2.
PD	
XX	
PD	06-JAN-2000.
XX	
PF	25-JUN-1999; 99WQ-US14398.
XX	
PR	26-JUN-1998; 98US-0105537.
XX	
PA	(MINU) UNIV MINNESOTA.
XX	
PI	Sherman DH, Liu H, Xue Y, Zhao L;
XX	
DR	WPI: 2000-160679/14.
XX	
DR	P-PsDB; AAY77177, AAY77178, AAY77199.
XX	
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT	synthesis of methymycin and pikromycin -
XX	
PS	Example 3; Figure 23; 438pp; English.
XX	
CC	The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the eryc gene cluster of Saccharopolyspora erythraea or streptomycetes antibiotics. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of polyketide synthetase may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer syntheses or to prepare biologically active agents, such as chemotherapeutics.
CC	Immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polyketides in plants. The present sequence represents a Streptomyces venezuelae ATCC 15439 DNA sequence, designated vcp ORF 1 in the specification, which actually contains 3 open reading frames, which encode proteins AAV77177-Y77178 and AAV77199. The vcp ORF 1 protein is defined in the specification as a PHA monomer synthase.
CC	
XX	
SO	Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;
	Query Match 50.9%; Score 51.4; DB 21; Length 15872;
	Best Local Similarity 72.0%; Pred. No. 0.0014;
	Matches 67; Conservative 0; Mismatches 26; Indels 0; Gaps 0
OY	9 GCTGCTGATCGCGGCCTGGAGACAACGGTTCGAGCGATCGCGCGGCTTCGCGCGC 68 Db 10606 GGTGTGATCTTCGCCGCGAGGAAGCGGTGCGCGAGCGCGCGCACACTCGCCGACG 10665
OY	69 CGGCTGGCGCAACAGCGGCTGCACTCTCGCA 101 Db 10666 AGGCGCGCGCACAGCGGCTTCGGGTGCGCA 10698

```

RESULT 5
AAF24892
ID AAF24892 standard; DNA: 20394 BP.
XX
AC AAF24892;
XX
DT 20-APR-2001 (first entry)
XX
DE pimarinin biosynthesis associated polyketide synthase gene.
XX
KM polyketide synthase; oxidative modification; metabolite; antibiotic;
RW anticancer; pimarinin; ss.
XX
OS Streptomyces natalensis.
XX
FH Key Location/Qualifiers
FT 1..20394
CDS /*tag= a
/product= "polyketide synthase"
PN WO20007722-A1.
XX
PD 21-DEC-2000.
XX
PE 14-JUN-2000; 2000WO-EP06227.
XX
PR 14-JUN-1999; 99EP-0201893.
XX
PA (STAM ) DSM NV.
XX
PI Martin JF, Aparicio JF, Colina AJ;
DR WPI: 2001-080693/09.
DR P-PSDB; AAB31558.
XX
PT New polynucleotides encoding enzymes involved in the biosynthesis of
PT pimarinin, useful for modifying the biosynthesis of pimarinin and in
PT the synthesis of new compounds -
XX
PS Disclosure; page 53-80; 116pp; English.
XX
CC The present sequence encodes a polyketide synthase which is associated
CC with the biosynthesis of pimarinin. The polyketide synthase polypeptide
CC is useful for the oxidative modification of a methyl group of a suitable
CC compound, e.g. a bioactive compound including a secondary metabolite,
CC antibiotics and anticancer agents. Recombinant cells comprising the
CC gene are useful for the production of pimarinin. The polyketide synthase
CC polynucleotide may be over expressed in Streptomyces, leading to an
CC increase in the biosynthesis of pimarinin, as a source of primers for
CC amplification reaction and as probes.
XX
SQ Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;
QY Query Match 50.9%; Score 51.4; DB 22; Length 20394;
DB Best Local Similarity 69.3%; Pred. No. 0.0014;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0.
QY 1 CGGAGCAGGTGGTATCGCGGGCGTGAGACAAGCGGTGCAGGCGATCGGCGGGGTTTC 60
DB 11944 CCGTGTGGTGTCTGATCTCGGGGCCGAGGAGCGGCTGTGGCATCGGCGGCACTTC 12003
QY 61 GGGGGCGCGCGCTGCGCACCAAGGCGGTGATGTCTGCCA 101
DB 12004 GCGGGGGAGGGGCGCGCAAAACCACCGACTGCGGGTCTGCCA 12044
RESULT 6
AAF80413
ID AAF80413 standard; DNA: 43280 BP.
XX
AC AAF80413;

```



```

FT      /product- "NysL protein"
FT      complement (58786..58980)
FT      /tag- e
FT      /product- "NysM protein"
FT      /note- "CDS does not include start codon"
FT      complement (59045..60241)
FT      CDS
FT      /tag- f
FT      /product- "NysN protein"
FT      /note- "CDS does not include start codon"
FT      complement (60238..61296)
FT      CDS
FT      /tag- g
FT      /product- "NysD2 complete protein"
FT      120628..121308
FT      /tag- h
FT      /product- "NysR4 (Lcng) protein"
XX      MO200159126-A2.
XX      16-AUG-2001.
XX      PD
XX      PF      08-FEB-2001; 2001MO-GB00509.
XX      PR      08-FEB-2000; 2000GB-0002840.
XX      PR      10-APR-2000; 2000GB-0008786.
XX      PR      14-APR-2000; 2000GB-0009387.
XX      PA      (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX      PA      (SNTF) SINTEF STIETTESEN IND TEK FORSK.
XX      PA      (ALPH-) ALPHARMA AS.
XX      PA      (SINV-) SINVENT AS.
XX      PA      (DZIE/) DZIEGLEMSKA H.
XX      PA      (ZOTC/) ZOTCHEV S B.
XX      PA      (SEKU/) SEKUROVA O N.
XX      PA      (EJAE/) EJAERVIK E.
XX      PA      (BRAU/) BRAUTASET T.
XX      PA      (STRO/) STROM A R.
XX      PI      Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX      PI      Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX      DR      WPI: 2001-557614/62.
XX      DR      P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142,
XX      DR      AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX      DR      AAE10149, AAE10150.
XX      PT      New nystatin polyketide synthase polynucleotides and polypeptides,
XX      PT      useful as antibiotics and antifungals -
XX      PS      Claim 1: Page 188-254; 266pp; English.
XX      CC      The present invention relates to the cloning and sequencing of the gene
XX      CC      cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX      CC      involved in the biosynthesis of the macrocyclic antibiotic nystatin.
XX      CC      The nystatin PKS is useful as antifungal antibiotics. The present
XX      CC      sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX      SQ      Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

Query Match      43.0%; Score 43.4; DB 22; Length 125401;
Best Local Similarity 66.7%; Pred. No. 0.089; 31; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 31;

XX      9 GGTGCTGATCGCGGCGTGGAGCAAGCGGTGAGGCGATCGCGGCGGTTCGCGCGCG 68
XX      DB      65109 GGTGCTGATCGCGGCGGAGCGGAGCGGCGGTGAGGCGATCGCGCGCGCG 68
XX      OY      69 CGGCGTGGCGACCAAGCGGCTGCATCTTCGCA 101
XX      DB      65169 CGGCGGCGGACCAAGCGGCGGCTGCATCTTCGCA 65201
XX      RESULT 10
XX      AAD17185
XX      ID      AAD17185 standard; DNA: 27541 BP.

```

```

XX      AC      AAD17185;
XX      DT      29-NOV-2001 (first entry)
XX      DE      Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX      KW      Polyketide synthase; PKS; macrocyclic; nystatin; PKS gene cluster;
XX      KW      antifungal; antibiotic; nys2; ds.
XX      OS      Streptomyces noursei.
XX      FH      Key Location/Qualifiers
XX      FT      CDS complement (454..1191)
XX      FT      /tag- a
XX      FT      /product- "NysF protein"
XX      FT      /note- "CDS does not include start codon"
XX      FT      complement (1275..3092)
XX      FT      CDS
XX      FT      /tag- b
XX      FT      /product- "NysG protein"
XX      FT      complement (3070..4824)
XX      FT      CDS
XX      FT      /tag- c
XX      FT      /note- "CDS does not include start codon"
XX      FT      /product- "NysH protein"
XX      FT      5122..6156
XX      FT      /tag- d
XX      FT      /product- "NysD3 protein"
XX      FT      6338..27541
XX      FT      /tag- e
XX      FT      /product- "NysI partial protein"
XX      FT      /note- "CDS does not include stop codon"
XX      PD      MO200159126-A2.
XX      PF      16-AUG-2001.
XX      PR      08-FEB-2001; 2001MO-GB00509.
XX      PR      08-FEB-2000; 2000GB-0002840.
XX      PR      10-APR-2000; 2000GB-0008786.
XX      PR      14-APR-2000; 2000GB-0009387.
XX      PA      (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX      PA      (SNTF) SINTEF STIETTESEN IND TEK FORSK.
XX      PA      (ALPH-) ALPHARMA AS.
XX      PA      (SINV-) SINVENT AS.
XX      PA      (DZIE/) DZIEGLEMSKA H.
XX      PA      (ZOTC/) ZOTCHEV S B.
XX      PA      (SEKU/) SEKUROVA O N.
XX      PA      (EJAE/) EJAERVIK E.
XX      PA      (BRAU/) BRAUTASET T.
XX      PA      (STRO/) STROM A R.
XX      PI      Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX      PI      Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX      DR      WPI: 2001-557614/62.
XX      DR      P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX      PT      New nystatin polyketide synthase polynucleotides and polypeptides,
XX      PT      useful as antibiotics and antifungals -
XX      PS      Claim 2: Page 151-166; 266pp; English.
XX      CC      The present invention relates to the cloning and sequencing of the gene
XX      CC      cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX      CC      involved in the biosynthesis of the macrocyclic antibiotic nystatin.
XX      CC      The nystatin PKS is useful as antifungal antibiotics. The present
XX      CC      sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX      SQ      Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match      41.8%; Score 42.2; DB 22; Length 27541;

```

Best Local Similarity 65.3%; Pred. No. 0.18;
Matches 62; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 1 CGGAGCAGGTGATCCGGCGGCGTGAGCAAGCGGTGAGCGGAGTCCGGCGGAGTTC 60
DB 13838 CCGAGTCGGTGTGATCTCCGTGACGAGGAGCGCCGAAACGATCCGCCACAGTTC 13897

OY 61 GCGGCGCGGCGGTGCGGCACAGCGGTGCATGT 95
DB 13898 GCCGAAACGCGGCGGCGCAAGACCAAGCGGCTCGCGGT 13932

RESULT 11
AAH52062

ID AAH52062 standard; DNA; 4851 BP.

AC AAH52062;

DT 04-SEP-2001 (first entry)

DE Mycobacterium tuberculosis potential drug target gene SEQ ID 116.

KW Drug target; growth; organism viability; characterisation; ds.

OS Mycobacterium tuberculosis.

PN WO200135317-A1.

PD 17-MAY-2001.

PF 13-NOV-2000; 2000WO-US31152.

PR 12-NOV-1999; 99US-0165086.

PR 12-NOV-1999; 99US-0165124.

PR 01-FEB-2000; 2000US-0179531.

PA (REGC) UNITV CALIFORNIA.

PI Eisenberg D, Rotstein SH, Marcotte EM;

DR WPI: 2001-329193/34.

DR P-PSDB; AAG81211.

PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences -

PS Disclosure; Page 131-133; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.

XX Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;

Query Match 41.4%; Score 41.8; DB 22; Length 4851;

Best Local Similarity 63.4%; Pred. No. 0.24;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 1 CCGAGCAGGTGATCCGGCGGCGTGAGCAAGCGGTGAGCGGAGTCCGGCGGAGTTC 60
DB 712 CCGAATCGGTGTGATCTCCGTGACGAGGCGCGCAAAATGCGATTCGATCGGTTTC 771

OY 61 GCGGCGCGGCGGTGCGGCACAGCGGTGCATGTCTCGCA 101
DB 772 GCCGCGCAGGCTGCGGCGGTGACACAGTTGCGGTCTCGCA 812

RESULT 12
AAH79278

ID AAH79278 standard; DNA; 31422 BP.

AC AAH79278;

DT 04-DEC-2001 (first entry)

DE Streptomyces avermectilis coding sequences SEQ ID NO: 2.

KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; veterinary drug; pesticide; ds.

OS Streptomyces avermectilis.

FT Key Location/Qualifiers

FT CDS 1..14646

FT /*tag= a

FT /*product= "AAG65266"

FT /*tag= b

FT /*product= "AAG65267"

PN WO200162939-A1.

PD 30-AUG-2001.

PF 23-FEB-2001; 2001WO-JP01381.

PR 24-FEB-2000; 2000JP-0047405.

PA (KYOW) KYOMA HAKKO KOGYO KK.

PA (KITA) KITASATO INST.

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;

DR WPI: 2001-582053/65.

DR P-PSDB; AAG65266, AAG65267.

PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -

PS Disclosure; Page 103-149; 257pp; Japanese.

XX The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermectilis genome.

XX Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

Query Match 39.8%; Score 40.2; DB 22; Length 31422;

Best Local Similarity 64.5%; Pred. No. 0.52;
Matches 60; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 9 GGTGATCGGCGGCGGTGAGCAAGCGGTGAGCGGAGTCCGGCGGAGTTC 68
DB 10284 GGTGTGTTCTCCGCGCGGAGGACGAGGTGGGAAATGCGACTGTCCGCAAGCG 10343
OY 69 CCGCGTCCGACCAAGCGGCTGCATCTTCGCA 101

Db 10344 CGGGCGAGAGTCAAGCGCTGCGACCGGCA 10376

RESULT 13

AAA92302
ID AAA92302 standard; DNA: 31422 BP.

AC AAA92302;

DT 10-JAN-2001 (first entry)

DE S. avermiltillis avermectin aglycon synthase DNA avail SEQ ID NO:2.

KM Streptomyces avermiltillis; avermectin aglycon synthase; biosynthesis;

KW multifunctional enzyme; polyketide; avermectin; veterinary drug;

KW agrochemical; ds.

OS Streptomyces avermiltillis.

FT Key Location/Qualifiers

FT CDS 1..14646

FT CDS /note= "avermectin aglycon synthase protein"

FT CDS /tag= b

FT CDS /note= "avermectin aglycon synthase protein"

PN MO200050605-A1.

PD 31-AUG-2000.

PF 23-FEB-2000; 2000MO-JP01041.

PR 24-FEB-1999; 99JP-0046961.

XX (KITA) KITASATO INST.

XX Omura S, Ikeda H;

XX WPI: 2000-565458/52.

DR P-PSDB: AAB23751, AAB23752.

XX Avermectin aglycone synthase DNA and proteins encoded by all or part of

PT it for the production of avermectin and its derivatives for drug and

PT agrochemical use

XX Claim 2; Page 134-203; 314pp; Japanese.

XX The present sequence represents DNA which encodes avermectin aglycon

CC synthase proteins. Also described are: (1) polypeptides encoded by all

CC or part of the DNA; (2) expression vectors containing the DNA; (3) host

CC cells transformed by the vectors; (4) preparation of avermectin aglycon

CC by culture of the transformants; (5) preparation of avermectin aglycon

CC or its derivatives by culture of transformed avermectin-producing

CC microorganisms; and (6) oligonucleotides of 5-60 bases in length

CC containing sense or antisense sequences from the avermectin aglycon

CC synthase DNA. The enzymes are useful for the production of modified

CC forms of avermectin and of the intermediates in its biosynthesis, for

CC use as drugs, veterinary drugs and agrochemicals.

XX Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

Query Match 38.2%; Score 38.6; DB 21; Length 31422;

Best Local Similarity 63.4%; Pred. No. 1.2;

Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 9 GGTGATGTCGCGGCGCTGAGCAAGCGGTGACAGCGATCGCGGCGGTTCCGGCGCG 68

DB 10284 GGTGATGTCGCGGCGCTGAGCAAGCGGTGACAGCGATCGCGGCGGTTCCGGCGCG 10343

QY 69 CGGCGTGCAGCAAGCGCGCTGCTCTCTGCA 101

DB 10344 CGGCGCGAGAGTCAAGCGCTGCGACCGGCA 10376

RESULT 14

AAQ10212/c
ID AAQ10212 standard; DNA: 2721 BP.

AC AAQ10212;

DT 17-DEC-2001 (updated)

DT 27-MAR-1991 (first entry)

DE BamHI J-I fragment carrying sequences characteristic of productive

DE pseudorabies virus.

KW PRV; ss.

OS Pseudorabies virus.

PN USN7537855-N.

PD 18-DEC-1990.

PF 13-JUN-1990; 90US-0238940.

PR 13-JUN-1990; 90US-0537855.

XX (USDA) US AGRIC RES SERV.

XX Cheung AK;

XX WPI: 1991-021957/03.

XX Pseudo-rabies virus nucleotide sequences - used for producing

XX nucleic acid probes, antigens and antibodies for distinguishing

XX latent from productive infection

XX Disclosure; Page 21; 27pp; English.

XX The fragment carries sequences characteristic of the productive

CC pseudorabies viral genome, and may be used as a probe in diagnosis

CC of infection.

CC (Note: Revised entry submitted to correct the patent number format of

CC US Government-owned NTIS applications to prevent clashes with ongoing US

CC granted patent numbers. For further information please visit the Derwent

CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)

XX Sequence 2721 BP; 428 A; 1007 C; 1017 G; 269 T; 0 other;

Query Match 37.8%; Score 38.2; DB 12; Length 2721;

Best Local Similarity 61.6%; Pred. No. 1.7;

Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 2 CGGAGAGGTGATGTCGCGGCGCTGAGCAAGCGGTGACAGCGATCGCGGCGGTTCCG 61

DB 787 CGGAGAGGTGATGTCGCGGCGCTGAGCAAGCGGTGACAGCGATCGCGGCGGTTCCG 728

QY 62 CGGCGCGCGGCGGTGACAGCGGTGACAGCGGTGACAGCGGTGACAGCGGTGACAGCG 100

DB 727 CGGCGCGCGGCGGTGACAGCGGTGACAGCGGTGACAGCGGTGACAGCGGTGACAGCG 689

RESULT 15

AAQ10543/c
ID AAQ10543 standard; DNA: 2721 BP.

AC AAQ10543;

DT 17-DEC-2001 (updated)

DT 27-MAR-1991 (first entry)

DE BamHI J-I fragment carrying sequences characteristic of latent

DE pseudorabies virus.

KW . PRV; ss.
XX
OS Pseudorabies virus.
XX
PN USN7537855-N.
XX
PD 18-DEC-1990.
XX
PF 13-JUN-1990; 90US-0238940.
XX
PR 13-JUN-1990; 90US-0537855.
XX
PA (USDA) US AGRIC RES SERV.
XX
PI Cheung AK;
XX
DR WPI; 1991-021957/03.
XX
PT Pseudo-rabies nucleotide sequences - used for producing
PT nucleic acid probes, antigens and antibodies for distinguishing
PT latent from productive infection
XX
PS Disclosure; Page 22; 27pp; English.
XX
CC The fragment carries sequences characteristic of the latent
CC pseudorabies viral genome, and may be used as a probe in diagnosis
CC of infection.
CC (Note: Revised entry submitted to correct the patent number format of
CC US Government-owned NTIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ Sequence 2721 BP; 428 A; 1007 C; 1017 G; 269 T; 0 other;

Query Match 37.8%; Score 38.2; DB 12; Length 2721;
Best Local Similarity 61.6%; Pred. No. 1.7;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 2 CGGACGAGTGTGATCCGGCGGCGTGAGCAGCGGTGCAGCGCATCCGGCGGGGTTGC 61
DB 787 CGGCGGCGGCGGAGGTGGCGGCGGTGTGAGAGCGGCGGCGGCGGAGAGGCT 728

QY 62 CGGCGCGGCGGCGTGCGCACAGCGGCGTGATGTCGC 100
DB 727 CGGCGGCGGAGATGTCGCCGCTCCCTTCCTCCCGC 689

Search completed: November 5, 2002, 12:33:56
Job time : 242.215 secs

[illegible]

OTSTRMETTLDIKRPLMIGHRVQGANVFPAGATVLEMAISSGATLADGDPLQITDVV
LAELAFAGDAVAVLVQVVTTEQPSGRILFOLASRPGAGHAFRVARHARALVRTE
VPAGITLAVRARLQASPPAATVETLQMTQASRPAQVQAGLFRHARALVRTE
DAAGSAAYRLHPALDADCFQVYVGLFAGGAGATVFWPEVYVGLLRLPSELECHA
RVNNGROTPOGADFWVYVSSGAVVAEVSGLVAVNORLPGRVRRREDDWPLEMEHA
AAVCTAKYNAKRWMLLGGGGGIGALRLSMLEAGNAYVHAASNTSAAGVRLALKAF
DQATFAYVHLSLGGGELDPGLAGAGALDPRASADSPDLADLALVRCQSVLMTV
QALAGMFRDAPRMLLTGAGAVAGADVSTQATLIGRVIMAEHADLRCARVLDL
PTRPGEIGALLAEILDAAEVALRGECVARIVRQEPTRRGHIESCVPTDVT
IRADSTVLTGGLGLGVAGMLAERGAHGLVLRGSGAASVEGRAAVVALEAGAR
VTYKADVADRAOLERITREVTTSGMTRGVYHAGITIDGLMOTCARAPKVNAPK
VOGALHIALUTREAPLSTFVITASVGLISPGCSTYAAANTFLDALHHRRAQPLA
LSVDGLFAEVMAAQEDRGARLVSRKMRSLTPDEGLSALRLIESRVOVGCPV
PRLWELYPAAASSRMLSLVYAHASAGPAGDGLFGLRLAAAPASRSGLEBLR
AOISQVLRLPBKEIEVDALITSLGMSLIMGLERNRIEMGLITVPTILMTYPTVAA
LSGLHARACEAAPVESPHTTADSAVEIEEMSDOLTLIAKFRALT"

16134 .37907
/gene="epod"
16134 .37907
/gene="epod"
/note="type 1 polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/transl_table=1
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQNPLKQAIIIORELERLGLAQLERTEPIAIV
GICRFGAGADAPAFWELDAERDAPVQDPRRAVLVAPVAPVPHNARGLITEIDC
FDAFFGISPREARSLDQHRILLEVAEAGLEDDAIPRSDIGSTGTFVGAFTADYA
RTVARLPREEDVATSNMSTIAAGRLSYTLGGLPCLVDTAGSSILVALHLCRS
LRAGSDSLADAGVSTLSIPDMEAAARTOLSPRCRTFPAASNGVRCGCLV
LKRLSDAQRDSDIWAIRGSAINDGRSTGLTANVALEALARSAAVEAGV
DYETHGTSIDGPIEVEALRATVAPSGTRCVLAVKNTKNGHLLAAGAVNLIC
AALSTHERIRIPRNIRPLEGSALATPRVWPRDRRFRGVSFSGSGT
NAHYLEAPAVEMLPAPERSAELVLVSGSEGLDQALREHLDHPELGLDV
AFSLATTSAMTHRLAVTSREGILVLSVAOQTPAGARCLASSRGLALIFT
GOGAOTPGMRGLCAWPAFRAPEDRCVLTLDRELDRLREMAEAGSABSLIDOT
AFTOPALFAVEYALTAIWRMVGVEPELLVHSGISGLVACVAGVSLDGYALVARG
RLMGLISAGAMVSLGAPEAEVAAVAPHAAMVSTIAVNGPROVYIACVAVQALIA
GPAAGVTRKRLHSHAFHSPLEMPLEPFRGVAASVYTRRPSVLSVNSIKVITDE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 4,9e-10;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGAGCAGTGTATCCGGCGCGTGAGCAAGCGGTCCAGCGCGCGGGGTTTC 60
|||||
Db 18300 CCGGAGCAGTGTATCCGGCGCGTGAGCAAGCGGTCCAGCGCGGTTC 18359
|||||

QY 61 GCGGCGCGCGGTGCGCACCAAGCGGTGATGTCTGCCA 101
|||||
Db 18360 GCGGCGCGCGGTGCGCACCAAGCGGTGATGTCTGCCA 18400
|||||

RESULT 2
LOCUS AR172664 71989 bp DNA linear PAR 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Jullien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothilones C and D
JOURNAL Patent: us 6303342-A 2 16-OCT-2001;
FEATURES
source 1. 71989
/organism="unknown"
/protein_id="AAF62883.1"

BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN

Query Match 100.0%; Score 101; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 4,7e-10;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGAGCAGTGTATCCGGCGCGTGAGCAAGCGGTCCAGCGCGGGGTTTC 60
|||||
Db 18300 CCGGAGCAGTGTATCCGGCGCGTGAGCAAGCGGTCCAGCGGTTC 18359
|||||

QY 61 GCGGCGCGCGGTGCGCACCAAGCGGTGATGTCTGCCA 101
|||||
Db 18360 GCGGCGCGCGGTGCGCACCAAGCGGTGATGTCTGCCA 18400
|||||

RESULT 3
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain so ce90 epothilone biosynthesis gene
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangiineae; Polyangiaceae; Polyangium.
1 (bases 1 to 68750)
Molnar,I., Schupp,T., Ono,M., Ziklie,R., Milmanow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Strattan,A., Cyr,D.D.,
Goriach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum so ce90
Chem. Biol. 7 (2), 97-109 (2000)
20130945
MEDLINE
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
Molnar,I.
Direct Submission
Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agricultural Research Institute, Inc., 3034 Cornwallis Rd, P.O. Box
112357, Research Triangle Park, NC 27709, USA
LOCATION/Qualifiers
1. 68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:36"
<1. 1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDLAFARATSAVLDGHRATERVHVALEARGIEDRLAREH
LRIOGSPSPMCIGDLTVLELADIDPLASISFHHASLRHPDPTSDAMLVDPALV
RMLARAGAPLREYERERRARATLQALRMGAAPCPAPDLIPREDDNGLIPG
MSPEVAERLRASATPELACALLAWLSTGAPMSGYATPELBNLLIGGLPT
ATAASAPGTSEALRGARLPASWEVVSRSKSGQNTPEALMERLITVAMGADN
LSREFRAAIAEVNRLAOPAPFAAGSLAVAGVSSGRLSGLVTDALYSGDND
IVMEQPGISIVVLACTDPPELAPPLSQMLFVAHANAGTISKVLEGSPLIYARN
OARPSLIVHAGFAMVAVQAVWPDERGAPVYVOSTIMEFHPTRCLHBPASAFS
LACDEHLYMCELSAGRLERHPRHPRGASRAFYIGEHLIATWYPSLTINATHVL
MADPRRAILVQDKRTGTEPYLAETRRPRAHYVSEDDITALLGQDPSRDMVYEHIR
SGASTVADYORQMDRDPVNLNRGLFTTNDRLTIARS
complement(1900. 3171)
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to DP-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"

Best Local Similarity 99.0%; Pred. No. 9.3e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 60
|||||
Db 23912 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 23971
|||||

QY 61 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 101
|||||
Db 23972 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 24012
|||||

RESULT 4
LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Liyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9.3e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 60
|||||
Db 23912 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 23971
|||||

QY 61 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 101
|||||
Db 23972 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 24012
|||||

RESULT 5
LOCUS AR199551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Liyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9.3e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 60
|||||
Db 23972 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 24012
|||||

Db 23912 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 23971
QY 61 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 101
|||||
Db 23972 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 24012
|||||

RESULT 6
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Liyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9.3e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 60
|||||
Db 23912 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 23971
|||||

QY 61 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 101
|||||
Db 23972 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 24012
|||||

RESULT 7
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION AR199567
VERSION AR199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Liyon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355459-A 1 12-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 98.4%; Score 99.4; DB 6; Length 68750;
Best Local Similarity 99.0%; Pred. No. 9.3e-10;
Matches 100; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 60
|||||
Db 23912 CCGGACAGGTGTGATTCGGGGCGGTGGAGCAAGCGGTGCAGCGCATCCGGGGGTTTC 23971
|||||

QY 61 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 101
|||||
Db 23972 GCGGCGCGGGCGGTGGCAACAGCGGTGCATGTCTCGCA 24012
|||||

HAHSPLEMLADRFRAVADGMTAARIPVVSNTGRPATAEELCCAEYVNGHREA
VRFADGVGLRQGVYTFLELPDGSLSALAEASADDVLAPVLRKNPEAPALLTA
IARHAGCGPVMASAPAGTGTARWDLPTPYAHOHERMPSGGAANRGSGAGISAG
HPLICAEVLASGGRLTGRSLISSHPLADHYVLSVYVCTILMELYLRADAEVD
CAVDELTIAAPLVPRASGALIQVWVGEPEDEAGHRVSVHARIBEGVTILHAGAL
AAAEETVPEPTDIAIWPPOGAEHDAACCYERFADAGFAYGVFOGJRAAKLKEDIYAE
VALPEGTGDAHLHPALFDALHAALLEGECTDEAAVPEFNNVYTLAHATGASRYER
IRPEGTGTSIALVDYAGAVASVRSILAPITVAGOLQOTGRDLSLQVDMTTLHDER
ANSIALIKGDPGIIIDTSLIOPHADLDLDAAGVHDVTIAPLPTPTAGTVESVHAAT
GALAIIRSLADRRRAASLTVPTVNGAVSGTDLACASVNGLYRSALLBERPGEGLVDY
DYDQAEPLVPRALASDEPOVLVRGEGVLARLRADSSDVTYNDPSTVLTIGGTG
GLGRSVARHIVSEHVSRLILVSRGPAEVDALVAELCAVYATVACVOTDAVA
VADLVARHISAVHTAGVLDGVESLPERLSAVLRPKVADANMLHATGSLDIDA
FVFSVAGTFCGAGACNANAEFLDALYHRRAYGLPRVLANPMSODGDMGTGL
SDAYORLAROGMPPLTYEEGALFDALDAAGHAPVRLDALRACGEPOLIRG
LIRTRGRRRAAATGEGDMAAFAGRLTGSAAGREPVYIGARSQIACVLGAHEATEI
DODRAFLDGTSLTAVELRNRLGAVGTIGRLPTEVLRPTPAELVAHHAATIAEPPT
VGEPLALGELERMEKSFGLDITTEHNEHJAGRLEVLRAKMDALINDTAAAGHDSPTS
DEDPDFESASDEVPDLIDNEJLS"

BASE COUNT 2611 a 7486 c 7397 g 2900 t

ORIGIN

Query Match 50.9%; Score 51.4; DB 6; Length 20394;
Best Local Similarity 69.3%; Pred. No. 0.88;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 CCGGACGAGGTGGTGAATCGCGGCGGTGGACGAGCGGTGCAGCGCGGCGGCGGCTTC 60
Db 11944 CCGTCGTGGTGGTGAATCGCGGCGGTGGACGAGCGGTGCAGCGCGGCGGCACTTC 12003

QY 61 GCGGCGCGCGCGCGTGGCGCACCAAGCGCTGCATGCTCGCA 101
Db 12004 GCGGCGGAGGCGGCGCAAAACCACCGACTGCGGCTCGCA 12044

RESULT 14
SFU78289 43280 bp DNA linear BCT 13-AUG-1997
LOCUS Streptomyces fradiae tylactone synthase, starter module and modules
DEFINITION 1-7 (tylg) gene, complete cds.
ACCESSION U78289
VERSION 078289.1 GI:2317859
KEYWORDS
SOURCE Streptomyces fradiae.
ORGANISM Streptomyces fradiae.
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 43280)
Dehoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
TITLE Sequence of Streptomyces fradiae tylactone synthase gene tylg
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 43280)
Dehoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
TITLE Direct Submission
JOURNAL Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
Center, Indianapolis, IN 46285, USA
COMMENT
FEATURES
source
Location/Qualifiers
1. 43280
/organism="Streptomyces fradiae"
/db_xref="taxon:1906"
1. 43280
/gene="tylg"
816..14234
/gene="tylg"
/note="ORF1; putative"
/codon_start=-1
/transl_table=11
/product="tylactone synthase starter module and modules 1
& 2"
/protein_id="AAB66504.1"
/db_xref="GI:2317860"
/translation="MSSALRAVQNSNGYGDLMSTNTAQNQGDJEDVDGPSTHGGE


```

/translation="TRPDAREATEPYGALLAPLMPERSADREVEGTKMTSDMSQARFL
AHDAVEVLLTLAGQPLVLLLEDLQMADTASLDLRLESPRCQHPGLGYLTARFE
TESDATTERRIVSDVLRGPRTALRLCGLRQAVGLAEHNSGVDPPVEVLEHARSE
GNPYFVMOULLVGDARLRHPDADALLATPPGVRREALNORFALPEPVLRYRLC
AVTGAEVETDLRRATCDASVNAIGESATIRAGLIGEPHHPGQJLHFNALVOETLFED
ELPRDRRRRLHAKVADALYARRRGVGOEEIERVAHSHWSKASLPAAEVLPRLLAA
BHAEQAMAEBOVEIWLRRRAVHLVGLPPDDPATVLEQRLHLOLQVLTATGYGHPH
AETALTJRGARLSTVTHADEPDSVLMALCAASLYTGRYASROLGCLDRLSMOTREPY
AVTGAAYEGEITLYHRGMLPOLAELEHGERADHPANGHALATTPQHDPPVCSRY
NTEFWMLGHRO IADARRRELLSTREYSRPSDRSFALYDGVVNAWEGDPTFLAAS
DEGVVACGDHGLTYKAKMLGVKGMGLVHAGDRDGLAMHRTSLAELQSKRYTLHPL
HLGLGQAOHHNGQITEARITFRMLAAVEQRERERYLHPALPAVPLLEHLLGHGAE
ALADRPGRGY"
/gene="p1mk"
/2190..3566
/gene="p1mk"
/function="putative glucosyl transferase"
/translation="MESARPLTFVSLPESGLNBLVLAGELSHQGEDLMFATDEP
RNDVKRIAEGSPVEFASLGEYDSEMSAVTMSDEYVRETOPSRTKARRAVVHTYRP
GLQAEKFRRLAVIDEVOPLVMDICISGFVADAIARNIPYLVIPPLSPVLRNHT
HFAKSYTRGPVPVPTGLSRKMTLAQRVANELFKLTRAMPLNPLGLVYLAADNRRT
ELQPKASFMAIETHADLYLCLNSLELDYPTDIPKMRVGMANVPLPAPDDQLSR
WLDQSSVYVYGLGTYTRLTREQVSNVEARLRDRHQVLMKLRSEQOHLPPRESL
PGLNREVSEVPSQMDVLAPHVKEFTFGGNGEMGMFGKPLVYRPLWDCYDOAV
RGODGLSLTLDRPQTIQIDNVYDKLTIVLIGPSPYEKAERRAALMSRAGGRETAAGL
VLSLPALA"
complement(3662..9736)
/gene="p1ms4"
/complement(3662..9736)
/gene="p1ms4"
/function="polyketide synthase"
/codon_start=1
/translation="MNDKPNDKLTVDYLKWTNLDLHQTROBLREVESAKOEPILAVG
MSCRLPGVRSPEEFWMOLLKQGRDGIINTEPDRGMDLILAGDGSSTATAGCGTDA
ASFDAGFPGCISPREAVTMDPQORILLESWEALERAGIDPVSLKTEGCVPTGTCGID
TINTVNSREDIEGHSTGLSILISGRVSTFGLEGPAVTVDTACSSLSVLHAAQ
ALRVGECISLALAGVTALSTPMSFVFGSNQGLADGRCARASDADGTAMSEGVVL
YVERLSDARRNHPVYLAVRGSAVNOGDASNGLTAPSGSOQRVJROALAAAGLSAD
VDVVAHGHTGALDGPTEAOALATYGRDROBERPLILGYSKNTGHQOAAAGAAGLI
KMILLMRHGVLPRTLHVARPSHVDSSGALQLLEPDMPRKADSPMRAGVSSPAISG
TNAHITLQPPAAPEREDDEGPHTSAPVPLVSKASEBALTAQDLDRISALTDRSPLD
LGLSLATTRAHLEHRAVLLSEGGTEVARGAHEHTAGLAVLFSGQAQRGMKREL
YGRFVFAEBALDGLVLAHFDAGLRDVMFCEBGLDETGTOPALPAIEVALFPLSSLG
YRDPVACHSISEIAAHVAGVSLADACALAAARALMAGCAMAAYAAASGV
AARLVGVSIIVAAVNGPEAVVIAEGEREVYLRIAEFTGLGKTRRLPVSHAFSPISML
MEDFRVVAEGILSYEPARLAVNSLTGELATMEDYARSAEYRRRLREYRFRDPSAL
AAEGASVLELGGDGVUTRAMOHTIDGTATTYPALRKDSEETALLTALAOLHVAGV
VDMGSIIFAGTGARVLDLPTYPQHERWYRPRAVLSGLDVAGAGLRSAEHPRLCALITLA
ESGEVILTGRSLQAYPMLTJDTMGGSVLEPATGSLDLMARADVDGDRDDEFVLTML
PLYLAEEGTOYVULGPADENAGRTISIHARPDDDPQPMQHAGGVLTSGERLVD
GSTVMPQGAEAVDLEGEFYDTJGYPGFOGLRQSVMOCEGAAVEVALPQVAVDGAIF
GLHPYLDAVLOSRRMAGIGAAQDTPFLPRAMGVSILHAGSALVILARARYKKGEDPILS
VAVDAEGARVUSVEELVLRTOULTHTDAPLTRARGARBRALLSLWNSGPEVPVACVRC
VSLGADDELGVASVATLADLSDDEDVVAVPVSGVGDPAAAHETLARTLGLLDSWG
DEGTASASRLVEVTRNAVASDGEESVADAAVAAVMGLVRSQOSEPGRALVDLEDSG
BLAEVTLACGLPGLAAGDAQOFYRREGAVRGRLVESNEGAGVSPREMDAGTYLVAGTG
GGIGRELRHLHLYVERGVHLLVRSRGEASVGLRDEILAHGAEIIVRACNDVADRA
AVEAVYGVADANHLPLVITHTAGVLDGDCYTIASITPERLSAVLRPKVDAAMNHEATKD
IDLAFVLYSSISVGLSIGQCNVAAGVFIDALAGHRAGLCPQOSLAWKMTQCGG
MTATLSDADQMNASYGTAPLSPERGLALFLDAATVGSATHVLAVGRVSGPVAAQSPV
SMLRGLVGGARTAVNTGETAVSLADRLRAARPERQOLMDVVTRTEAAALVAGGSS
ONVADAREFEYELGDSLTISVELNRSLSTVGTGLSATVVFQSKNPTDLAARLYALDAA
OSALDGSASQAGFLPVAPETDLSLERMFLDALDSCVPEAQRMLSALGALRSFENTA
```

```

complement(9772..15198)
/gene="p1ms3"
/complement(9772..15198)
/gene="p1ms3"
/function="polyketide synthase"
/codon_start=1
/translation="MDQOQKLNDYIKRASADIRNSROVGELEIAAREPILAVTMSCR
FPGVARSPEELMSLADAGDGTGKSPARQWELLAEDEGESGGLHDAAEFASFEFG
ISPRBALMDPOORILLSSMSAEFRAAIDPAVAVGTGTGHEVGMPEYRGPPDDV
OGFALTGTTSYISGRLAYFGAVGAPVLAQVREARLAEALHLEHTSRPGCLAPLDVA
YSIATARSGLERBAFTAAQPGAGALATLADGPPAGGLVODTARTSKLAFIAGQ
GSOBGMGBRELAAAPRVATLDELVLGHEGDELEPRLPVLPAAGCEPALLDTOYQ
KOPALFATEVALYRLAESFGITTPDLASHSISGEITAAHVAGVFTTLADAALVYLRGL
MOALPEGGAMVLETEEEVVLPLEDRDRULSTIAVNPAAVNVIAGEDADVTAAVEHF
DALGRTTRLRVSHAFHSPLEMLAERTAVSRLTPAPVPIPVASGLTGLTAAVEOL
TSPDYLHARVAVEFADALRLARARRRHRLPMQPDVLSLMAQOCCRGHGRHRR
LRHARVVRGREAEVTLPAALAGLHTRGVAVVOWMEPYFGGTGARRADLPTVAFQRRRYV
BLBDLPLPATLAEGPGPRLICVSTPTANGVHEYARLAASFGEHRVYSLPLVGFAT
GERLPEATREAVYIAESALASGPNPLVDEGSGSAGCFATYALALNTENTJPEAV
VLDLTLRHHONESIDYAGLARHREHVEDVSPVMTNSRLSTAAARMMGMNOLEVRH
TTPVLTJIRAKKEIKFICGTDTGITYGEDHGSVDVRSVADAHFSVRDAPETATIVE
WLDLSGNA"
complement(15210..43733)
/gene="p1ms2"
/complement(15210..43733)
/gene="p1ms2"
/function="polyketide synthase"
/codon_start=1
/translation="MDOQKLDYIKRASADIRNSROVGELEIAAREPILAVTMSCR
FPGVARSPEELMSLADAGDGTGKSPARQWELLAEDEGESGGLHDAAEFASFEFG
ISPRBALMDPOORILLSSMSAEFRAAIDPAVAVGTGTGHEVGMPEYRGPPDDV
OGFALTGTTSYISGRLAYFGAVGAPVLAQVREARLAEALHLEHTSRPGCLAPLDVA
YSIATARSGLERBAFTAAQPGAGALATLADGPPAGGLVODTARTSKLAFIAGQ
GSOBGMGBRELAAAPRVATLDELVLGHEGDELEPRLPVLPAAGCEPALLDTOYQ
KOPALFATEVALYRLAESFGITTPDLASHSISGEITAAHVAGVFTTLADAALVYLRGL
MOALPEGGAMVLETEEEVVLPLEDRDRULSTIAVNPAAVNVIAGEDADVTAAVEHF
DALGRTTRLRVSHAFHSPLEMLAERTAVSRLTPAPVPIPVASGLTGLTAAVEOL
TSPDYLHARVAVEFADALRLARARRRHRLPMQPDVLSLMAQOCCRGHGRHRR
LRHARVVRGREAEVTLPAALAGLHTRGVAVVOWMEPYFGGTGARRADLPTVAFQRRRYV
```

PKSLPATGCDYVRANGUGAHHPLLTAAVSVANSDDLITGRLSRTHPWLAADNAVGT
VLLPGTAFLELAVRAGDEAGCGVREELTLAAPLVLPREGGVQVWVWSPDASGRRAV
SVHSRDPGPELPMTOHAGTLVAGEHHAGFDATVMPADATPLDLDFYARMAADTGF
GYGPLEFGGLRAAMRSCDDVYAEVALPDSGAASTESSAAGLHPALDHALHAAAFVD
LGEDARGLPEFSMODVTLHASGASTVRVKLTPDGGDAVALAVADTGTGAPVASTGSLVL
RTLDPDQIDAHSLVYRDAIPELRMSERKOTAHHRAPASAVLGPDLFGLERALAPTDLT
VTIAPGPDAAHPHADVYLPVPAAGTTEDDVVTSVHARTAHVLERLRTWLTDERTASRL
VEVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGREFGLLDLPCTTEPATBPLRLA
LALSDEPOTAVRGTAVALARLAQAAPQAATEWDEPGTVLGTAGTGLGRVLAHRLV
AERGVRLILVSRGADALGAGELMAELAHGAOVTLAACDLADRAVRRLSGVPAP
HPVTAAVHSAGSLDDGVYIASLTPERTSAVLRPRKADAVWHLHEATROLDLAAFYVESL
SGTVGAAGCGNVAANAFIDALQLRRAOGLPCLSLAWGPMAPTAGTGGTLTDDDLR
LARAGTPTALTEEGIALFETATATDAVLLPTRIDLSVLKQSEIPLPWRGLITTPAR
RSVVSASETAVTLVQQLSFLAETDRREVLDLVAGQVAAVLGHAGTADIDPRRPLREL
GFDSLTAVELRNRLSATGCLRTAATVIFDHPVTDALAHLLDGLMGSSEAAAAADAPRT

Query Match 50.9%: Score 51.4; DB 1; Length 84985;
Best Local Similarity 69.3%; Pred. No. 0.66;
Matches 70; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 CCGGAGCAGGTGTCGATCGCGCGGTGAC CAGCGGTGCAGCCGATCGCGGGGTTTC 60
||| | ||||| ||||| ||| | ||||| ||| ||||| |||
Db 69209 CCGTCGTGGTCTGTGATCTCGGGCGGACGACGGGTCTGGCCATCGCGGCACTTC 69150

QY 61 GCGGCGCGCGCGGTGCGCACCAAGCGGCTGCATCTCGCA 101
||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 69149 GCGGGGGAGGGGCGCAAAACCCAGCTCGGGTCTCGCA 69109

Search completed: November 5, 2002, 12:08:15
Job time : 630.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 : Search time 1003.2 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065

Perfect score: 91

Sequence: 1 ccgacgcgcacacggatggc.....tcgacccgggttcacgcgc 91

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	37.4	577	17	A0162259 mgx00012K
2	30.6	33.6	1160	12	BG844853
3	30.2	33.2	817	12	BG343906
4	29.8	32.7	553	12	BE725325
5	29.8	32.7	603	13	BE725325
6	29.8	32.7	668	13	BE725325

7	29.6	32.5	560	17	A2935138
8	29.6	32.5	656	17	A2935404
9	29.6	32.5	658	14	B0765891
10	29.4	32.3	488	14	BM691860
11	29.4	32.3	623	17	BH807613
12	29	31.9	398	10	BE739356
13	29	31.9	523	10	AV618984
14	29	31.9	549	10	BE219908
15	28.6	31.4	224	9	AI920329
16	28.6	31.4	510	12	BE697901
17	28.6	31.4	561	9	A0148928
18	28.4	31.2	387	12	BC343971
19	28.4	31.2	401	17	AQ090925
20	28.2	31.0	219	14	BQ322897
21	28.2	31.0	605	10	AM286258
22	28.2	31.0	679	10	BB594172
23	28.2	31.0	874	12	BE254077
24	28.2	31.0	881	17	CNS03K1K
25	28.2	31.0	1026	17	CNS04LVH
26	28	30.8	421	13	B1264863
27	28	30.8	594	10	BE504110
28	27.8	30.5	469	10	BE583753
29	27.8	30.5	512	9	A1588339
30	27.8	30.5	523	13	BE723385
31	27.8	30.5	530	10	BE583754
32	27.8	30.5	557	13	BE723387
33	27.8	30.5	601	13	BE723386
34	27.8	30.5	609	10	BE583577
35	27.8	30.5	646	10	BE583752
36	27.8	30.5	965	9	AL532447
37	27.8	30.5	1057	17	CNS03BPC
38	27.8	30.5	1064	14	BQ948399
39	27.8	30.5	1174	14	BQ876942
40	27.6	30.3	224	17	BE221657
41	27.6	30.3	585	14	BQ034876
42	27.6	30.3	666	10	BB653198
43	27.6	30.3	1274	14	BQ065081
44	27.4	30.1	268	14	BQ759911
45	27.4	30.1	299	12	BE462534

ALIGNMENTS

RESULT 1
LOCUS A0162259 577 bp DNA linear GSS 09-SEP-1998
DEFINITION mgx0012K23r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
clone mgx0012K23r, DNA sequence.
ACCESSION A0162259
VERSION A0162259.1 GI:3558660
KEYWORDS
SOURCE
ORGANISM
GSS.
Magnaporthe grisea.
Magnaporthe grisea.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes; incertae sedis; Magnaportheaceae; Magnaporthe.

REFERENCE

1 (bases 1 to 577)
Yu, F., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R.,
Phillips, K., Sasnowski, M., Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
Genome

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1998)
Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293

Email: rdeane@clemson.edu
Seq primer: GGAAACGCTATGACCAG
Class: BAC ends
High quality sequence stop: 337.
Location/Qualifiers

FEATURES

```

source
1. .577
/organism="Magnaporthe oryzae"
/strain="70-15"
/db_xref="taxon:148305"
/clone="mgx0012k23r"
/clone_1kb="CGI Rice Blast BAC Library"
/tissue_type="Prictoplasts"
/lab_host="E. coli DH10B"
/note="Vector: pEACWICH; Site 1: HindIII; Site 2: HindIII;
Rice blast is one of the most devastating fungal diseases
of rice world wide. It is a filamentous ascomycete with
a haploid genome (n=7) of approximately 40 Mbp. Rice
blast is an important model fungal pathogen for studying
numerous aspects of the fungal-host interaction. In
order to facilitate genome wide analysis, a BAC library
containing 9216 clones with an average insert size of 130
kbp was constructed. This library represents greater
than 25x genome coverage. High density colony filters
are available upon request."
121 a 148 c 148 g 160 t

```

[illegible]

RESULT 2	
BG844853/c	
LOCUS	BG844853
DEFINITION	1160 bp mRNA linear EST 29-MAY-2001
ACCESSION	U04008A08.y2 C. reinhardtii CC-1690, normalized, Lambda zap II
VERSION	BG844853
KEYWORDS	Chlamydomonas reinhardtii cDNA, mRNA sequence.
SOURCE	BG844853.1 GI:14226037
ORGANISM	EST. Chlamydomonas reinhardtii. Chlamydomonas reinhardtii

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1160)	Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McInerney,J.P., Silflow,C., Stern,D. and Surzycki,R.	Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analysing Gene Function and Regulation in Vascular Plants; project phase 2	Unpublished (2000)	Contact: Charles Hauser

FEATURES
source
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers
1. 1160

```

clone.lib="C. reinhardtii CC-1690, normalized, lambda zap
II"
/notes="Vector: pBluescript II SK-; Site.1: EcoRI; Site.2:
XhoI; This library, constructed by John Davies and Jeffrey
Medemont, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in

```

ambient levels of CO₂ and HS medium bubbled with 5% CO₂. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

[illegible]

RESULT 3	
RG343906	
LOCUS	817 bp mRNA linear EST 22-Oct-2001
DEFINITION	HVSMBg0007B09f Hordeum vulgare pre-anthesis spike EST library
	HVCDNA0008 (white to yellow anther) Hordeum vulgare cDNA clone
	HVSMBg0007B09f, mRNA sequence.
ACCESSION	RG343906
VERSION	RG343906.1 GI:13156235
KEYWORDS	EST.
SOURCE	Hordeum vulgare.
ORGANISM	Hordeum vulgare.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Poideae 1 (bases 1 to 817)	Wing, R., Close, T. J., Kleinholz, A., Wise, R., Begum, D., Frisch, D., Yu, X., Healy, D., Palmer, M., Rambo, T., Simmons, J., Choi, D. W., Fenton, R. D., Close, S. J., Oates, R. and Malm, D.	Development of a genetically and physically anchored EST resource for barley genomics: Morex pre-anthesis spike cDNA library	Unpublished (2001)	Contact: Wing RA

100 Jordan Hall, Clemson, SC 29634, USA
Tel.: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total hg bases = 418
Seq primer: AATTAACCTCCTACTAAGG
High quality sequence start: 6
High quality sequence stop: 798.
Location/Qualifiers
I. 817

notes="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI; plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ close, TJ close). Whole spike with awns trimmed were collected at white, green and yellow anther stages (Fenton). Total RNA was prepared from each pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one

Primary unamplified cDNA library was made, and 1 million pfu were *in vivo* excised to give phagescript SK(-) cDNA phagids. These steps were performed in the T7 Close Lab (Choi) at the University of California, Riverside. Phagids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing) Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinbous A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. *Barley Genetics Newsletter* 3:1:29-30. <http://wheat.pw.usda.gov/gspages/bgn/31/covar.html>."

BASE COUNT	148 a	286 c	263 g	120 t
ORIGIN				

Query Match	33.2%	Score 30.2;	DB 12;	Length 817;
Best Local Similarity	69.5%;	Pred. No. 1.4e+02;		
Matches 41;	Conservative 0;	Mismatches 18;	Indels 0;	Gaps 0;

Oy 11 ACGGGATGGCCACCGCATCTGGGGCTGATCCGGGGCTCGGCCATCAACCATGATGCCC 69
 ||||| | ||||| ||||| | ||||| ||| | | ||||| ||
 Db 653 ACGGGCGAGACCCCATCTGGGGCTTAACCGGGGAGGGGCTCCGGCTTGATGACCC 711

RESULT 4	
LOCUS	BE725325/c
DEFINITION	BE725325 553 bp mRNA linear EST 14-SEP-2000 894082D06.x1 C. reinhardtii CC-1690, normalized, lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

VERSION	BE725325.1	GI:10126605
KEYWORDS	EST.	
SOURCE	Chlamydomonas reinhardt	

Source	Organism
Chlamydomonas reinhardtii.	Chlamydomonas reinhardtii
Chlamydomonas reinhardtii	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 553)

AUTHORS	TITLE
Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefevre, P., McDermott, J. P., Sillow, C., Stern, D. and Surzycki, R.	Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2

JOURNAL
COMMENT

Unpublished (2000)
Contact: Charles Hauser

Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES	Location/Qualifiers
source	1. .553

```

/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, lambda zap
II"

```

/note="Veccor, pbluscript II SK-; site_1: EcoRI; site_2: XhoI; This library, constructed by John Davies and Jeffrey Modermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the

ambient levels of CO₂ and H₂ medium bubbled with 5% CO₂. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambdaZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda

BASE COUNT	99 a	202 c	148 g	104 t
ORIGIN	ZAP clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."			

Query Match	32.7%	Score 29.8	DB 12	length 553
Best Local Similarity	58.4%	Pred. No. 1.6e+02		
Matches 52	Conservative 0	Mismatches 37	Indels 0	Gaps 0

QY 2 CGAGCGGCACGGGATGGCGACCCTCTGTGGCGCTCATCCGGGGCTCGGCCATTCAACCA 61
| | |||| | | ||| ||| | | |||| |||| ||| | | ||||
Db 418 CAAGCGGCACGGCAGAAGACGGCTCGCGAGTCGACGCGCTTCCGCGGCTACGCGCTGAACCT 35

QY	62	TGATGGCCGGTGCACCGGTTGACCGCGC	90
Db	358	TGCCCGCGCGGCGACGGCATGCCCAAGC	330

RESULT 5			
B1530795/c			
LOCUS	B1530795	603 bp	mRNA
DEFINITION	10241108611.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II		
	Chlamydomonas reinhardtii cDNA, mRNA sequence.		

ACCESSION	B1530795
VERSION	B1530795.1
KEYWORDS	GI:15371365
SOURCE	EST.
	<i>Chlamydomonas reinhardtii</i>

ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.

AUTHORS	TITLE
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J. P., Shrago, J., Silflow, C. and Stern, D.	Analyses of the Chlamydomonas reinhardtii Genome: A Model,

JOURNAL
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1024b
unpublished (2001)
COMMENT
Contact: Charles Hauser

COMMENT

Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chausere@duke.edu

FEATURES	Location/Qualifiers
source	1. .603

```

/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, lambda Zap
II"

```

*Xho*I-*Not*I; pBluescript II SK-*+*, Site 1: EcoRI, Site 2: KcoRI, Site 3: *Xho*I). This library, constructed by John Davies and Jeffrey Kocemoltz, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO₂ and HS medium bubbled with 5% CO₂. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambdaZAP II (Stratagene) in the EcoRI (5') and *Xho*I (3') sites. pBluescript II SK- plasmids were excised from the lambdaZAP clones by superinfection with Exsist1 (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT	104 a	220 c	167 g	112 t
ORIGIN				

Query Match	32.7%	Score 29.8	DB 13	length 603
Best Local Similarity	58.4%	Pred No. 1.6e+02		
Matches 52; Conservative	0	Mismatches 37	Indels 0	Gaps 0

2 CGACGCGCAACGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61

Db 418 CAAGGCGCCAGCGGCAAGAGCCCTCGAGTGCAGCCTCTCGCGGCTACGCGTGAACCT 359
 Oy 62 TGATGGCCGGTTCGACCGGTTGACCCGGC 90
 Db 358 TGCCCGCGCGCGCGGCGCATGCCCAAGC 330

RESULT 6
 B1530912/c
 LOCUS B1530912 668 bp mRNA linear EST 29-AUG-2001
 DEFINITION 1024109909.x2 C. reinhardtii CC-1690, normalized, lambda zap II
 ACCESSION B1530912
 VERSION B1530912
 KEYWORDS GI:15371486
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE
 Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
 1 (bases 1 to 668)
 P., McDermott, J.P., Shreger, J., Silflow, C. and Stern, D.
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants. Project: 1024p
 Unpublished (2001)
 CONTACT: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu

TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1..668
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mc+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, lambda zap
 II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI, Site 2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP (acetate-containing) medium in the
 light, TAP medium in the dark, HS (minimal) medium in
 ambient levels of CO2 and HS medium bubbled with 5% CO2.
 polyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 zap clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method A described
 in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 116 a 236 c 184 g 132 t
 ORIGIN

Query Match 32.7%; Score 29.8; DB 13; Length 668;
 Best Local Similarity 58.4%; Pred. No. 1.7e+02;
 Matches 52; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Oy 2 CGACGGCAACGGGATGCGACCGCATCTTGGCGTGATCCGGCGCTCAACCA 61
 Db 418 CAAGGCGCGGCAAGAGCCCTCGAGTGCAGCCTCTCGCGGCTACGCGTGAACCT 359
 Oy 62 TGATGGCCGGTTCGACCGGTTGACCCGGC 90
 Db 358 TGCCCGCGCGCGCGGCGCATGCCCAAGC 330

RESULT 7
 A2935138
 LOCUS A2935138 560 bp DNA linear GSS 24-APR-2001

DEFINITION BJ_Ba0003K18r B. japonicum BAC library Bradyrhizobium japonicum
 genomic, DNA sequence.
 ACCESSION A2935138
 VERSION A2935138.1 GI:13777370
 KEYWORDS GSS.
 SOURCE Bradyrhizobium japonicum.
 ORGANISM Bradyrhizobium japonicum.
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE
 Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Golcochea
 1 (bases 1 to 560)
 J.T., Stacey, G., Sadowsky, M.J. and Wing, R.A.
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
 genome

TITLE
 JOURNAL
 MEDLINE
 COMMENT

FEATURES
 source
 1..560
 /organism="Bradyrhizobium japonicum"
 /strain="USD110"
 /db_xref="taxon:375"
 /clone_lib="B. japonicum BAC library"
 /lab_host="E. coli"
 /note="Vector: pIndigo536; Site 1: HindIII"
 Location/Qualifiers

BASE COUNT 117 a 164 c 172 g 107 t
 ORIGIN

Query Match 32.5%; Score 29.6; DB 17; Length 560;
 Best Local Similarity 59.5%; Pred. No. 1.8e+02;
 Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Oy 5 CGGCGCAACGGGATGCGACCGCATCTGCGCGTGATCCGGCGCTCAACCATCA 64
 Db 342 CGACAAATGACATGCGCGCCCAACAAGAGCGTTCAGGCGGTGTAACGATCT 401
 Oy 65 TGGCGGTGCGACCGGTTGACCGC 88
 Db 402 TGCGCTCGCCACAAGCGCGCCG 425

RESULT 8
 A2935404
 LOCUS A2935404 656 bp DNA linear GSS 24-APR-2001
 DEFINITION BJ_Ba0003008f B. japonicum BAC library Bradyrhizobium japonicum
 genomic, DNA sequence.
 ACCESSION A2935404
 VERSION A2935404.1 GI:13778084
 KEYWORDS GSS.
 SOURCE Bradyrhizobium japonicum.
 ORGANISM Bradyrhizobium japonicum.
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE
 Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Golcochea
 1 (bases 1 to 656)
 J.T., Stacey, G., Sadowsky, M.J. and Wing, R.A.
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
 genome

TITLE
 JOURNAL
 MEDLINE
 COMMENT

CONTACT: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288

[illegible]

DB	460	TGTCGGGGCTGGCGGGCC	478
RESULT 12			
LOCUS	BB739356/c		
DEFINITION	BB739356 RIKEN full-length enriched, 6 days neonate spleen Mus		
ACCESSION	BB739356		
VERSION	BB739356.1		
KEYWORDS	GI:16142361		
SOURCE	EST.		
ORGANISM	mouse mouse.		
REFERENCE	Mus musculus		
AUTHORS	Enkaiyotia, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 398)		
	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayata,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Komno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishii,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takeku-Akaiiri,S., Tanaka,T., Tomaru,A., Toya,T., Wachihi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. (2001)		
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Science Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayasu,N., Sugahara,Y., Shibata,K., Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Itawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,T., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.		
FEATURES			
SOURCE	Location/Qualifiers		
	1..398		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="F430113C10"		
	/clone_1fb="RIKEN full-length enriched, 6 days neonate spleen"		
	/tissue_type="spleen"		
	/dev_stage="6 days neonate"		
	/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI. This clone is among a rearrayed set of 15,247 clones from 11 embryonic cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All		

source libraries are cloned unidirectionally with Oligo(dT)
)-Not primers. References include: (1) Genome-wide
expression profiling of mid-gestation placenta and embryo
using a 15,000 mouse developmental cDNA microarray. 2000,
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132. (2)
Large-scale cDNA analysis reveals phased gene expression
patterns during preimplantation mouse development, 2000,
Development, 127: 1737-1749; (3) Genome-wide mapping of
unselected transcripts from extraembryonic tissue of
7.5-day mouse embryos reveals enrichment in the t-complex
and under-representation on the X chromosome, 1998, Hum
Mol Genet 7: 1967-1978."

BASE COUNT 62 a 144 c 115 g 77 t

ORIGIN

Query Match 31.9%; Score 29; DB 10; Length 398;
Best Local Similarity 58.8%; Pred. No. 2.4e+02;
Matches 50; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 3 GACGGCGAAGGATGGGACCGCATCTGGGGCTGATCCGGGCTCGCCATCAACCAT 62
Db 172 GACGGCGAAGGATGGGACCGCATCTGGGGCTGATCCGGGCTCGCCATCAACCAT 113

QY 63 GATGGCCGCTGACCGCGGTTGACCG 87
Db 112 AGAGCGCGTGCCTCCCGGCTTTCGCC 88

RESULT 13
AV618984 523 bp mRNA linear EST 15-DEC-2000
LOCUS AV618984 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LC002c01_r 5', mRNA sequence.
ACCESSION AV618984.1 GI:10768159
VERSION AV618984
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonadales.
REFERENCE 1 (bases 1 to 523)
AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..523
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone_lib="LC002c01_r"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

BASE COUNT 82 a 118 c 212 g 111 t

ORIGIN

Query Match 31.9%; Score 29; DB 10; Length 523;
Best Local Similarity 61.0%; Pred. No. 2.6e+02;
Matches 47; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 8 GCAAGCGATGGCGACCATCTGGGCGCTGATCCGGGCTCGCCATCAACCATGATGG 67
LOCUS A1920329 523 bp mRNA linear EST 29-JUL-1999

Db 446 GCATCGCATGCACAAGCAGAAGTGCTGCTGATGGCGGCGCTTGGCCGACACATGACGG 505
QY 68 CCGGTCGACCGGGTTGA 84
Db 506 CTGCTGGGGCGCGGGGA 522

RESULT 14
BE219908/c 549 bp mRNA linear EST 03-JUL-2000
LOCUS BE219908 h64d03.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3178181 3',
DEFINITION mRNA sequence.
ACCESSION BE219908
VERSION BE219908.1 GI:8907226
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 549)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.lnl.gov
Possible reversed clone: polyA not found
Seq primer: -40UP from Glbco
High quality sequence stop: 485.
Location/Qualifiers
1..549
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3178181"
/clone_lib="NCI CGAP Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP Lu5 was prepared, and ss circles were
made in vitro. Following BAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonids
141920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 133 c 149 g 117 t 1 others

ORIGIN

Query Match 31.9%; Score 29; DB 10; Length 549;
Best Local Similarity 61.0%; Pred. No. 2.7e+02;
Matches 47; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 2 GCAGCGCAAGGATGGCGACCGCATCTGGGCGCTGATCCGGGCGTCCGCATCAACCA 61
Db 78 CGCCCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 19

QY 62 TGATGGCGGTCGACCG 78
Db 18 TGATGGCGGCGCGCGCG 2

RESULT 15
A1920329 224 bp mRNA linear EST 29-JUL-1999

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 25.8918 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065

Perfect score: 91

Sequence: 1 ccgacgcgcgaacggatgacgc.....tcgaccgggttgaccgcgc 91

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	100.0	68750	3	US-09-335-409-1
2	91	100.0	68750	4	US-09-568-102-1
3	91	100.0	68750	4	US-09-567-969-1
4	91	100.0	68750	4	US-09-568-480-1
5	91	100.0	68750	4	US-09-568-486-1
6	91	100.0	68750	4	US-09-568-472-1
7	91	100.0	68750	4	US-09-567-899-1
8	91	100.0	71989	4	US-09-443-501A-2
9	45.2	49.7	4403765	4	US-09-103-840A-2
10	45.2	49.7	4411529	4	US-09-103-840A-1
11	44.6	49.0	33529	2	US-09-144-085-3
12	44.6	49.0	43280	2	US-08-804-227C-1
13	44.6	49.0	80161	3	US-09-036-987A-1
14	44.6	49.0	80161	4	US-09-370-700-1
15	43	47.3	643	4	US-08-861-774E-53
16	43	47.3	20235	3	US-07-642-734C-3
17	43	47.3	20235	3	US-08-439-009A-3
18	42	46.2	507	4	US-09-060-756-660
19	42	46.2	28958	1	US-08-258-261B-6
20	42	46.2	28958	1	US-08-456-837-6
21	42	46.2	28958	1	US-08-457-646A-6
22	42	46.2	28958	1	US-08-457-646A-6
23	42	46.2	28958	1	US-08-458-076A-6
24	42	46.2	28958	1	US-08-764-233A-4
25	42	46.2	28958	1	US-08-457-335A-6
26	42	46.2	28958	1	US-08-729-214-6
27	42	46.2	28958	3	US-09-028-934-6

28	42	46.2	49377	1	US-08-764-233A-1	Sequence 1, Appli
29	41.4	45.5	1462	4	US-09-434-288-4	Sequence 4, Appli
30	41.4	45.5	11220	4	US-09-105-537-32	Sequence 32, Appli
31	41.4	45.5	15872	4	US-09-105-537-1	Sequence 1, Appli
32	41.4	45.5	36778	4	US-09-105-537-5	Sequence 5, Appli
33	41.4	45.5	38506	3	US-09-320-878-19	Sequence 19, Appli
34	41.4	45.5	50937	4	US-09-428-517-1	Sequence 1, Appli
35	39.8	43.7	751	3	US-09-010-809-4	Sequence 4, Appli
36	39.8	43.7	4041	4	US-09-105-537-36	Sequence 36, Appli
37	39.8	43.7	44377	2	US-08-804-227C-7	Sequence 7, Appli
38	39.8	43.7	44377	2	US-08-804-198-1	Sequence 1, Appli
39	38.2	42.0	423	4	US-09-144-085-5	Sequence 5, Appli
40	38.2	42.0	754	4	US-08-861-774E-47	Sequence 47, Appli
41	38.2	42.0	13842	4	US-09-105-537-30	Sequence 30, Appli
42	36.8	40.4	643	4	US-08-861-774E-31	Sequence 31, Appli
43	36.6	40.2	13987	2	US-08-804-227C-13	Sequence 13, Appli
44	35	38.5	8035	4	US-09-215-694-29	Sequence 29, Appli
45	35	38.5	31328	4	US-09-215-694-19	Sequence 19, Appli

ALIGNMENTS

```
RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match          100.0%; Score 91; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred.No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGACCGCGACGAGGATGCGCGACCGCATCTGGGCGCTGATCCGGGCGCTGGCCATCAACC 60
    |||||||
DB 22587 CCGACCGCGACGAGGATGCGCGACCGCATCTGGGCGCTGATCCGGGCGCTGGCCATCAACC 22646
    |||||||

QY 61 ATGATGCGCGATCGACCGGATGACCGCGCC 91
    |||||||
DB 22647 ATGATGCGCGATCGACCGGATGACCGCGCC 22677
    |||||||

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
```

;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
DB 22587 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

QY 61 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 91
DB 22647 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 22677

RESULT 3
US-09-567-969-1
;; Sequence 1, Application US/09567969
;; Patent No. 6355457
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/567,969
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
DB 22587 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

QY 61 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 91
DB 22647 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 22677

RESULT 4
US-09-568-480-1
;; Sequence 1, Application US/09568480
;; Patent No. 6353458
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,480
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
DB 22587 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

QY 61 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 91
DB 22647 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 22677

RESULT 5
US-09-568-486-1
;; Sequence 1, Application US/09568486
;; Patent No. 6355459
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan
;; APPLICANT: Zirkle, Ross
;; APPLICANT: Cyr, Devon
;; APPLICANT: Goerlach, Joern
;; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;; FILE REFERENCE: 4-30582A
;; CURRENT APPLICATION NUMBER: US/09/568,486
;; CURRENT FILING DATE: 2000-05-10
;; PRIOR APPLICATION NUMBER: 09/335,409
;; PRIOR FILING DATE: 1999-06-17
;; NUMBER OF SEQ ID NOS: 30
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 68750
;; TYPE: DNA
;; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
DB 22587 CCGAGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

QY 61 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 91
DB 22647 ATGATGGCCGGTGCACCGCGTTGACCGCGCC 22677

RESULT 6
US-09-568-472-1
;; Sequence 1, Application US/09568472
;; Patent No. 6358719
;; GENERAL INFORMATION:
;; APPLICANT: Schupp, Thomas
;; APPLICANT: Ligon, James
;; APPLICANT: Molnar, Istvan

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCGAGCGCGAAGGGATGGCGACCGCATCTGGCGCTGTATCCGGGCTCGGCCATCAACC 60
|||
Db 22587 CCGAGCGCGAAGGGATGGCGACCGCATCTGGCGCTGTATCCGGGCTCGGCCATCAACC 22646
OY 61 ATGATGCCCGGTGACCGGGTTGACCGCGCC 91
|||
Db 22647 ATGATGCCCGGTGACCGGGTTGACCGCGCC 22677

RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match 100.0%; Score 91; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCGAGCGCGAAGGGATGGCGACCGCATCTGGCGCTGTATCCGGGCTCGGCCATCAACC 60
|||
Db 22587 CCGAGCGCGAAGGGATGGCGACCGCATCTGGCGCTGTATCCGGGCTCGGCCATCAACC 22646
OY 61 ATGATGCCCGGTGACCGGGTTGACCGCGCC 91
|||
Db 22647 ATGATGCCCGGTGACCGGGTTGACCGCGCC 22677

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:

; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct.
US-09-443-501A-2

Query Match 100.0%; Score 91; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCGAGCGCGAAGGGATGGCGACCGCATCTGGCGCTGTATCCGGGCTCGGCCATCAACC 60
|||
Db 16975 CCGAGCGCGAAGGGATGGCGACCGCATCTGGCGCTGTATCCGGGCTCGGCCATCAACC 17034
OY 61 ATGATGCCCGGTGACCGGGTTGACCGCGCC 91
|||
Db 17035 ATGATGCCCGGTGACCGGGTTGACCGCGCC 17065

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 49.7%; Score 45.2; DB 4; Length 4403765;
Best Local Similarity 68.9%; Pred. No. 6.4e-05;
Matches 62; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 2 CCGAGCGCGAAGGGATGGCGACCGCATCTGGCGCTGTATCCGGGCTCGGCCATCAACCA 61


```
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32
```

```
Query Match 45.5%; Score 41.4; DB 10; Length 11220;
Best Local Similarity 65.9%; Pred. No. 0.00013;
Matches 60; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
OY 1 CCGACGGCAAGGGATGGGACCGCATCTGCGGCTGATCCGGGGCTCGGCATCAACC 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 851 CCGACGGCGCTCCGACAGGGCCATCGATCCGCGCTGTCCGGGACCGCGCTCAACC 910
OY 61 ATGATGCGCGGTGACCGGGTTGACCGCGCC 91
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 911 AGGACGGCGCGACGACGGCGCTCAGCGCTC 941
```

```
RESULT 3
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1
```

```
Query Match 45.5%; Score 41.4; DB 10; Length 15872;
Best Local Similarity 65.9%; Pred. No. 0.00014;
Matches 60; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
OY 1 CCGACGGCAAGGGATGGGACCGCATCTGCGGCTGATCCGGGGCTCGGCATCAACC 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3843 CCGACGGCGCGGCAAGGGCCACCGCTGCTCGCGGTATCGGGGACGCGGTCAACC 3902
OY 61 ATGATGCGCGGTGACCGGGTTGACCGCGCC 91
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3903 AGGACGGCGCGCTCCACAGGGCGCTCAGCGCC 3933
```

```
RESULT 4
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
```

```
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5
```

```
Query Match 45.5%; Score 41.4; DB 10; Length 36778;
Best Local Similarity 65.9%; Pred. No. 0.00014;
Matches 60; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
OY 1 CCGACGGCAAGGGATGGGACCGCATCTGCGGCTGATCCGGGGCTCGGCATCAACC 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16538 CCGACGGCGCTCCGACAGGGCCATCGATCTCGCGCTGTCCGGGACCGCGCTCAACC 16597
OY 61 ATGATGCGCGGTGACCGGGTTGACCGCGCC 91
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16598 AGGACGGCGCGACGACGGCGCTCAGCGCTC 16628
```

```
RESULT 5
US-09-861-289-36
; Sequence 36, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 4041
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-36
```

```
Query Match 43.7%; Score 39.8; DB 10; Length 4041;
Best Local Similarity 64.8%; Pred. No. 0.00039;
Matches 59; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
OY 1 CCGACGGCAAGGGATGGGACCGCATCTGCGGCTGATCCGGGGCTCGGCATCAACC 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 CCGACGGCGGCGCAAGGGCCACCGGCTCTGCGCTGTCCGGGACGCGCATCAACC 907
OY 61 ATGATGCGCGGTGACCGGGTTGACCGCGCC 91
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 908 AGGACGGCGCGACGACGGCGCTCAGCGCTC 938
```

```
RESULT 6
US-09-924-256A-47
; Sequence 47, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
```

```
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 754
; TYPE: DNA
; ORGANISM: Parmelia sulcata
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (753)
; OTHER INFORMATION: n=a, c, g or t
US-09-924-256A-47
```

```
Query Match 42.0%; Score 38.2; DB 10; Length 754;
Best Local Similarity 63.7%; Pred. No. 0.0011;
Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
```

```
OY 1 CCGAGCGCGGATGGGCGACCGCATCTGGCGGCTGATCGGGGCGTGGCGATCAACC 60
DB 581 CGGAGCGCTTACGCGATACGACACTGCTCGTCACTGATGCGGCGCTTCTGTCACAC 640
OY 61 ATGATGCGCGGTGACCGGGGTGACCGCGCC 91
DB 641 AGGATGGCAAGTCTCCCGGATACACACACCC 671
```

```
RESULT 7
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30
```

```
Query Match 42.0%; Score 38.2; DB 10; Length 13842;
Best Local Similarity 63.7%; Pred. No. 0.0013;
Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
```

```
OY 1 CCGAGCGCGGATGGGCGACCGCATCTGGCGGCTGATCGGGGCGTGGCGATCAACC 60
DB 8555 CGGAGCGCGGCGGCGCAAGGTCACCGGCTCTCGCGGCTACCGCGGTCAACC 8614
OY 61 ATGATGCGCGGTGACCGGGGTGACCGCGCC 91
DB 8615 AGGAGCGCGCGCAAGCGGCTGACCGCGCC 8645
```

```
RESULT 8
US-09-924-256A-31
; Sequence 31, Application US/09924256A
```

```
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Leptogium corniculatum
US-09-924-256A-31
```

```
Query Match 40.4%; Score 36.8; DB 10; Length 643;
Best Local Similarity 63.6%; Pred. No. 0.0029;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
OY 4 ACCGCGAAGGATGGCGACCGCATCTGGCGGCTGATCGGGGCGTGGCGATCAACCATG 63
DB 474 AGCGGAGACACATGCGCATTAATTAATGACACATCAAGGGTCAAGCTCGAATCATG 533
OY 64 ATGCGCGGTGACCGGGGTGACCGCGCC 91
DB 534 GTGAGACGTCCGCGCGGCTCACCGGTAC 561
```

```
RESULT 9
US-09-775-938A-29
; Sequence 29, Application US/09775938A
; Patent No. US20020081665A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Haygood, M.
; APPLICANT: Davidson, S.K.
; APPLICANT: Allen, S.W.
; APPLICANT: Hildebrand, M.
; TITLE OF INVENTION: Bryostatins, Bryopyrans and Polyketides: Compositions and Meth
; FILE REFERENCE: 1133.010US1
; CURRENT APPLICATION NUMBER: US/09/775,938A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US00/21326
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/147,283
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 6000
; TYPE: DNA
; ORGANISM: Endobugula sertula
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (386)...(388)
; OTHER INFORMATION: TAG may represent a transposase open reading
; OTHER INFORMATION: frame.
; NAME/KEY: misc_feature
; LOCATION: (444)...(449)
; OTHER INFORMATION: TTGAAA may be a possible -35 trascription control
; OTHER INFORMATION: sequence.
; NAME/KEY: misc_feature
; LOCATION: (458)...(463)
; OTHER INFORMATION: GATTAAT may be a possible -10 trascription control
; OTHER INFORMATION: sequence.
; NAME/KEY: misc_feature
; LOCATION: (474)...(502)
; OTHER INFORMATION: ATCAATATAAA and TTTTATATGAT are inverted repeats.
```

NAME/KEY: misc_feature
LOCATION: (576)...(583)
OTHER INFORMATION: TGAGCAAT may be a possible SD sequence.
NAME/KEY: misc_feature
LOCATION: (565)...(567)
OTHER INFORMATION: ATG encoding M is presumptive start of rks Open
OTHER INFORMATION: reading frame.
NAME/KEY: misc_feature
LOCATION: (589)...(591)
OTHER INFORMATION: GTG encoding V is possible alternative start of
OTHER INFORMATION: rks Open reading frame.
US-09-775-938A-29

Query Match 40.2%; Score 36.6; DB 10; Length 6000;
Best Local Similarity 62.8%; Pred. No. 0.004;
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 CCGAGCGCAACGGGATGGGACCGCATCTGGGCGTGAATCCGGGGCTCGGCATCAACC 60
DB 2627 CGGATGGCGTGGCGTGAATGATCCCATTCGTGCAATGATACGGGGCTGGGCTGTGAATC 2686
QY 61 ATGATGGCGGTCGACCGCGTTCACCGCC 91
DB 2687 AGGATGGTACAGTAAATGATTTACGGGCC 2717

RESULT 10
US-09-924-256A-61
Sequence 61, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Walters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
TITLE OF INVENTION: BIOACTIVE MOLECULES
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924, 256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 61
LENGTH: 649
TYPE: DNA
ORGANISM: Pseudocyclophellaria anthrapiis
US-09-924-256A-61

Query Match 38.0%; Score 34.6; DB 10; Length 649;
Best Local Similarity 61.8%; Pred. No. 0.014;
Matches 55; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 3 GAGCGCAACGGGATGGGACCGCATCTGGGCGTGAATCCGGGGCTCGGCATCAACC 62
DB 479 GATGCTATTTGAGATGAGACTGTATAGAGCATATCCGGGCACTGGGCTCAACTCC 538
QY 63 GATGGCGGTCGACCGCGTTCACCGCC 91
DB 539 GATGGCAAGACCCCTGGCATTAACCTTGCC 567

RESULT 11
US-09-861-289-34
Sequence 34, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding metilmycin and pikromycin

FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861, 289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105, 537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 4689
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-34

Query Match 36.7%; Score 33.4; DB 10; Length 4689;
Best Local Similarity 60.4%; Pred. No. 0.039;
Matches 55; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 CCGAGCGCAACGGGATGGGACCGCATCTGGGCGTGAATCCGGGGCTCGGCATCAACC 60
DB 854 CGGAGCGCCCGCCGCAACGGACCGCATCTCGGCTGTGCGGCGGACGCGGTCACCC 913
QY 61 ATGATGGCGGTCGACCGCGTTCACCGCC 91
DB 914 AGGAGCGCGCCAGCACAGCGGCTTCACGGCTCC 944

RESULT 12
US-09-924-256A-45
Sequence 45, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Walters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
TITLE OF INVENTION: BIOACTIVE MOLECULES
FILE REFERENCE: 9993-006
CURRENT APPLICATION NUMBER: US/09/924, 256A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 08/861,774
PRIOR FILING DATE: 2001-04-13
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 45
LENGTH: 655
TYPE: DNA
ORGANISM: Parmelia sulcata
US-09-924-256A-45

Query Match 35.6%; Score 32.4; DB 10; Length 655;
Best Local Similarity 60.0%; Pred. No. 0.069;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 2 CGAGCGCAACGGGATGGGACCGCATCTGGGCGTGAATCCGGGGCTCGGCATCAACA 61
DB 472 CGCGCGGTCGGCGGAGGAGACCATTCATATGTCATACGGGACGCGGCTCAACA 531
QY 62 TGATGGCGGTCGACCGCGTTCACCGCC 91
DB 532 GGAGGGTGGCAGACCGGCTCTACCATGCC 561

RESULT 13
US-09-924-256A-51
Sequence 51, Application US/09924256A
Patent No. US20020127659A1
GENERAL INFORMATION:
APPLICANT: Walters, Barbara
APPLICANT: Miao, Vivian
APPLICANT: Ho, Yap
APPLICANT: Tong, Seow
TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR

```
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Peltigera neopolydactyla
US-09-924-256A-51
```

```
Query Match          35.6%; Score 32.4; DB 10; Length 703;
Best Local Similarity 60.0%; Pred. No. 0.069; Indels 0; Gaps 0;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
QY 2 CGACCGCAACGGATGGCGACCGCATCTGGGCGCTGATCCGGGCTCGGCCATCAACCA 61
DB 472 CGCTGCATTCCTGATGGAGATCATATCGATGCAATTCGCGGAGACAGAAATTACCA 531
QY 62 TGATGGCGGCTGCACCGGGTTGACCGCGCC 91
DB 532 GGATGGCAGAACGCGCGAATTACCATGCC 561
```

RESULT 14

```
US-09-924-256A-41
; Sequence 41, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Leptogium corniculatum
US-09-924-256A-41
```

```
Query Match          34.9%; Score 31.8; DB 10; Length 637;
Best Local Similarity 59.3%; Pred. No. 0.11;
Matches 54; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 1 CGAGCGCAACGGATGGCGACCGCATCTGGGCGCTGATCCGGGCTCGGCCATCAACC 60
DB 465 CCGATGCCGAAAGCGGTGTAATATCTGTGTGATCGAGGCTGGGGGTGAATC 524
QY 61 ATGATGGCCGCTGCACCGGGTTGACCGCGCC 91
DB 525 AAGACGGCAAGACCAAGTGGAATTACAGCACC 555
```

RESULT 15

```
US-09-924-256A-69
; Sequence 69, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
```

```
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 658
; TYPE: DNA
; ORGANISM: Usnea florida
US-09-924-256A-69
```

```
Query Match          34.1%; Score 31; DB 10; Length 658;
Best Local Similarity 58.4%; Pred. No. 0.19;
Matches 52; Conservative 1; Mismatches 36; Indels 0; Gaps 0;
```

```
QY 3 GACGGCAACGGGATGGGCGACCGCATCTGGGCGCTGATCCGGGCTCGGCCATCAACCAT 62
DB 473 GACGCTATTGCCGATGASATGCATCGATGCAATTCGAGACAGAGGTCCAAACCA 532
QY 63 GATGGCGGCTGCACCGGGTTGACCGCGCC 91
DB 533 GACGGCATTCAAATGTTATCAAGGTGCC 561
```

```
Search completed: November 5, 2002, 22:59:08
Job time : 88.2088 secs
```


FT	/product= "Orf 5 protein (AAV58584) "
PD	7610..11875
CDS	/*tag- f
FT	/product= "Type I polyketide synthase, EPOS A
FT	(AAV58573) "
FT	11872..116104
FT	/*tag- g
FT	/product= "Non-ribosomal peptide synthetase, EPOS P
FT	(AAV58574) "
CDS	16251..21749
FT	/*tag- h
FT	/product= "Type I polyketide synthase, EPOS B
FT	(AAV58575) "
CDS	21746..43519
FT	/*tag- i
FT	/product= "Type I polyketide synthase, EPOS C
FT	(AAV58576) "
FT	43524..54920
FT	/*tag- j
FT	/product= "Type I polyketide synthase, EPOS D
FT	(AAV58577) "
CDS	54935..62254
FT	/*tag- k
FT	/product= "Type I polyketide synthase, EPOS E
FT	(AAV58578) "
CDS	62369..63628
FT	/*tag- l
FT	/product= "Cytochrome P450 oxygenase homologue, EPOS F
FT	(AAV58579) "
CDS	63779..64333
FT	/*tag- m
FT	/product= "Orf 6 protein (AAV58585) "
FT	complement (63853..64290)
FT	/*tag- n
FT	/product= "Orf 7 protein (AAV58586) "
FT	64353..64920
CDS	/*tag- o
FT	/product= "Orf 8 protein (AAV58587) "
FT	complement (64287..64727)
FT	/*tag- p
FT	/product= "Orf 9 protein (AAV58588) "
FT	65053..65767
CDS	/*tag- q
FT	/product= "Orf 10 protein"
FT	complement (65008..65874)
FT	/*tag- r
FT	/product= "Orf 11 protein (AAV58590) "
FT	complement (65871..66338)
CDS	/*tag- s
FT	/product= "Orf 12 protein (AAV58591) "
FT	66667..67137
FT	/*tag- t
FT	/product= "Orf 13 protein (AAV58592) "
FT	67334..68251
CDS	/*tag- u
FT	/product= "Orf 14 protein (AAV58593) "
FT	68346..68750
FT	/partial
FT	/*tag- v
FT	/product= "Partial Orf 15 protein (AAV58594) "
FT	/note= "No termination codon given in the specification"
XX	
PX	W09366028-A2.
PD	23-DEC-1999.
XX	
PE	16-JUN-1999;
XX	99WO-EP04171.
PR	18-JUN-1998;
PR	24-SEP-1998;
PR	05-FEB-1999;
XX	98US-0099504.
XX	98US-0101631.
XX	99US-0118906.
PA	(NOVS) NOVARTIS AG.

PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PI	Schupp T, Liigon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX	WPI; 2000-097741/08.
DR	P-F5DB: AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR	AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR	AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
XX	AAY58592, AAY58593, AAY58594.
PT	New isolated epoethione synthase genes, used for the recombinant
PT	production of epoethione for use in cancer therapy -
XX	Claim 14; Page 87-104; 174pp; English.
PS	
CC	This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC	comprising 22 open reading frames (ORFs) and includes genes encoding
CC	proteins involved in the biosynthesis of epoethiones. Epoethiones A and
CC	B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC	starter unit; polyketides being synthesised from two-carbon building
CC	blocks, the beta-carbon of which always carries a keto group. Each round
CC	of two-carbon addition is carried out by a complex of enzymes known as
CC	the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC	EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC	the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
CC	and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC	formation. EPO F (AAY58579) is an epoethione macrolactone oxidase, and
CC	the proteins Orf 3 (AAY58582) and Orif4 (AAY58593) are thought to be
CC	involved in transport. Epoethiones mimic the biological activity of
CC	taxol, and may be substituted for taxol in cancer chemotherapeutic
CC	compositions. Epoethiones exhibit a much lower drop in potency against a
CC	multiply drug-resistant cell line compared with taxol, and are
CC	considerably less efficiently exported from such cells by the multidrug
CC	resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC	epoethiones as anticancer agents, they are problematical to produce on a
CC	large scale. Epoethiones are too complex for industrial scale chemical
CC	synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC	poor yields of epoethiones. The nucleic acids of the invention may be
CC	used for the recombinant production of epoethiones in a heterologous host
CC	that is more amenable to fermentation.
XX	
XX	
SO	Sequence 68750 BP: 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
	Query Match 100.0%; Score 91; DB 21; Length 68750;
	Best Local Similarity 100.0%; Pred. No. 1,2e-18;
	Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 CCGAGCGCGACGCGGATGGCGGCGCATTTGGGCGCTGATCCGGGGGCTCGGCATCAACC 60
DB	22587 CCGAGCGCGACGCGGATGGCGGCGCATTTGGGCGCTGATCCGGGGGCTCGGCATCAACC 22646
QY	61 ATGATGGCGCGGTGCAGCCGGGTTGACCGGCGCC 91
DB	22647 ATGATGGCGCGGTGCAGCCGGGTTGACCGGCGCC 22677
RESULT 2	
ID	AAA29349 standard: DNA; 71989 BP.
AC	AAA29349;
XX	
DT	12-SEP-2000 (first entry)
DE	Sorangium cellulosum epoethione polyketide synthase operon genomic DNA.
XX	
XX	Epoethione: polyketide synthase; eposA; eposB; eposC; eposD; eposE; eposF;
XX	eposG; eposH; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
XX	tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX	
OS	Sorangium cellulosum.
Key	Location/Qualifiers
TH	


```
FT CDS 3..992
FT /tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /tag= c
FT /label= epoc_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548
FT /tag= d
FT /note= "encodes ketide synthase (KS-Y) of the loading
FT domain"
FT misc_RNA 3621..4661
FT /tag= e
FT /note= "encodes acyl transferase (AT) of the loading
FT domain"
FT misc_RNA 4917..5610
FT /tag= f
FT /note= "encodes enoyl reductase (ER) of the loading
FT domain, potentially involved in formation of the
FT thiazole moiety"
FT misc_RNA 5856..6155
FT /tag= g
FT /note= "encodes acyl carrier protein (ACP) of the loading
FT domain"
FT CDS 6260..10493
FT /tag= h
FT /label= epob_gene
FT /note= "encodes module 1, the NRPS module"
FT misc_RNA 2031..3548
FT /tag= i
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 2031..3548
FT /tag= j
FT /note= "encodes condensation domain C2 of the NRPS
FT module"
FT misc_RNA 6861..6887
FT /tag= k
FT /note= "encodes heterocyclization signature sequence"
FT misc_RNA 6861..6887
FT /tag= l
FT /note= "encodes condensation domain C4 of the NRPS
FT module"
FT misc_RNA 7358..7366
FT /tag= m
FT /note= "encodes condensation domain C7 (partial) of the
FT NRPS module"
FT misc_RNA 7898..7921
FT /tag= n
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 7898..7921
FT /tag= o
FT /note= "encodes adenylation domain A1 of the NRPS module"
FT misc_RNA 8261..8308
FT /tag= p
FT /note= "encodes adenylation domain A3 of the NRPS module"
FT misc_RNA 8411..8422
FT /tag= q
FT /note= "encodes adenylation domain A4 of the NRPS module"
FT misc_RNA 8861..8905
FT /tag= r
FT /note= "encodes adenylation domain A6 of the NRPS module"
FT misc_RNA 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS
FT module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS
FT module"
FT CDS 10639..16137
FT /tag= x
FT /label= epoc_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AR2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epob_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AR3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AR4"
FT misc_RNA 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AR5"
FT misc_RNA 27966..28574
FT /tag= ao
FT /note= "encodes DHS"
FT misc_RNA 29433..30287
FT /tag= ap
```

```
FT /note= "encodes ER5"
FT misc_RNA
FT 30321..30869
FT /tag= aq
FT /note= "encodes KR5"
FT misc_RNA
FT 31077..31373
FT /tag= at
FT /note= "encodes AC15"
FT misc_RNA
FT 31440..32807
FT /tag= as
FT /note= "encodes KS6"
FT misc_RNA
FT 33018..34067
FT /tag= at
FT /note= "encodes AT6"
FT misc_RNA
FT 34107..34676
FT /tag= au
FT /note= "encodes DH6"
FT misc_RNA
FT 35760..36641
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA
FT 36705..37256
FT /tag= av
FT /note= "encodes KR5"
FT misc_RNA
FT 37470..37769
FT /tag= ax
FT /note= "encodes AC26"
FT CDS
FT 37912..49308
FT /tag= ay
FT /label= "epoe gene"
FT /note= "encodes moJules 7 and 8"
FT misc_RNA
FT 38014..39375
FT /tag= az
FT /note= "encodes KS7"
FT misc_RNA
FT 39589..40626
FT /tag= ba
FT /note= "encodes AT7"
FT misc_RNA
FT 41341..41922
FT /tag= db
FT /note= "encodes KR7"
FT misc_RNA
FT 42181..42423

Query Match 100.0%; Score 91; DB 21; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.2e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGACGGCGCAACGGATGCGACCGCATCTGGGCGCTGATCCGGGGCTCGCCATCAACC 60
DB 16975 CCGACGGCGCAACGGATGCGACCGCATCTGGGCGCTGATCCGGGGCTCGCCATCAACC 17034
QY 61 ATGATGGCGGCTCGACCGGGTTGACCGCGCC 91
DB 17035 ATGATGGCGGCTCGACCGGGTTGACCGCGCC 17065

RESULT 3
AADI17185 standard; DNA: 27541 BP.
AC AADI17185;
XX
XX 29-NOV-2001 (first entry)
XX Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
XX Polyketide synthase; PKS; macrolicide; nystatin; PKS gene cluster;
KM antifungal; antibiotic; nys2; ds.
XX
XX Streptomyces noursei.
XX
XX Key location/Qualifiers
FH complement (454..1191)
FT /tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
```

```
FT CDS
FT complement (1275..3092)
FT /tag= b
FT /product= "NysG protein"
FT CDS
FT complement (3070..4824)
FT /tag= c
FT /note= "CDS does not include start codon"
FT CDS
FT 5122..6156
FT /tag= d
FT /product= "NysD3 protein"
FT CDS
FT 6138..27541
FT /tag= e
FT /product= "NysI partial protein"
FT /note= "CDS does not include stop codon"

W0200159126-A2.
16-AUG-2001.
08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNIISK NATURVITENSKAPELIGE.
XX (SWF) SINTEF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIELEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (EJAE/) FJAEERVIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI: 2001-557614/62.
DR P-PSDB: AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolicide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
XX Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match 56.0%; Score 51; DB 22; Length 27541;
Best Local Similarity 72.5%; Pred. No. 2.1e-06;
Matches 66; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 CCGACGGCGCAACGGATGCGACCGCATCTGGGCGCTGATCCGGGGCTCGCCATCAACC 60
DB 17145 CCGACGGCGCAACGGATGCGACCGCATCTGGGCGCTGATCCGGGGCTCGCCATCAACC 17204
QY 61 ATGATGGCGGCTCGACCGGGTTGACCGCGCC 91
DB 17205 AGGACGGCGGCTCGACCGGGTTGACCGCGCC 17235

RESULT 4
AADI17186 standard; DNA: 125401 BP.
ID AADI17186
AC AADI17186;
```



```
OY 1 CCGAGCGCAACGGGATGCGACCGCATCTGCGGCGCTGATCCGGGCGTCGCCATCAACC 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 830 CCGATGCGACGCGGTCGACGCGCATCTGATCTGCGGCGCATGCGGATCGGCCGTCACAC 889
OY 61 ATGATGCGCGGTCGACCGCGGTTGACCGCGCC 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 890 ACGGCGCGCGCGACGACGCGCGCTCAGCGCGCC 920

RESULT 8
AAF90037
ID AAF90037 standard; DNA: 6462 BP.
XX
AC AAF90037;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic;
    type I polyketide synthase; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..6462
FT /tag= a
FT /product= "type I polyketide synthase"
FT /transl_except= "(1..3, aa: Met)"
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PE 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET ) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
DR P-PSDB; AAB83974.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Clatm 35; Page 307-309; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC a liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC represent open reading frames (ORFs) of the coding strand of cosmid
CC a26g1, and encode type I polyketide synthases.
XX
SQ Sequence 6462 BP; 1073 A; 2067 C; 2135 G; 1187 T; 0 other;
```

```
Query Match 52.5%; Score 47.8; DB 22; Length 6462;
Best Local Similarity 70.3%; Pred. No. 1.7e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
```

```
OY 1 CCGAGCGCAACGGGATGCGACCGCATCTGCGGCGCTGATCCGGGCGTCGCCATCAACC 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 857 CCGATGCGACTGCGCGCGCGAGACCGGGGTGTGGGCCGTGTGCGCGCTCCGCGGTCAATC 916
OY 61 ATGATGCGCGGTCGACCGCGGTTGACCGCGCC 91
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 917 AGGATGCGCGGTCGCGCGCGGCGTCACCGCTCC 947

RESULT 9
AAF90033
ID AAF90033 standard; DNA: 34071 BP.
XX
AC AAF90033;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS Synthetic.
XX
FH WO200140497-A2.
FT CDS 1..34071
FT /tag= a
FT /product= "type I polyketide synthase"
FT /transl_except= "(1..3, aa: Met)"
XX
PN 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET ) AVENTIS PHARMA SA.
XX
PI Jeanin P, Pernodet J, Guerneau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI: 2001-374849/39.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics
XX
PS Example 14; Page 289-300; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC a liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand), which encodes different
CC polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;
```

```
Query Match 52.5%; Score 47.8; DB 22; Length 34071;
Best Local Similarity 70.3%; Pred. No. 2e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
```

```

DB 19140 AGGATGGCCGCTCGGCCGCGCTCACCGCTCC 19170
RESULT 10
AAF90032/C
ID AAF90032 standard; DNA: 42717 BP.
AC AAF90032;
DT 06-AUG-2001 (first entry)
DE Nucleotide sequence of cosmid a26g1 (non-coding strand).
NM Metabolic pathway operon: polyketide; polyketide antibiotic; ss.
XX Synthetic.
XX WO200140497-A2.
XX 07-JUN-2001.
XX 07-JUN-2001.
XX 27-NOV-2000; 2000WO-FR03311.
XX 29-NOV-1999; 99FR-0015032.
XX 07-JUN-2000; 2000US-0209800.
XX (AVER ) AVENTIS PHARMA SA.
XX Jeanin P, Pernodet J, Guerin M, Simonet P, Courtols S;
XX Capellano C, Francou F, Raynal A, Ball M, Sezonov G, Tiphille K;
XX Frostegard A;
XX WPI: 2001-374849/39.
XX
XX Collection of nucleic acids from environmental samples, useful for
XX identifying e.g. genes encoding polyketide synthases and derived
XX antibiotics -
XX
XX Example 14; Page 274-288; 356pp; French.
XX
XX The specification describes a method for the preparation of a collection
XX of nucleic acids from organisms in a soil sample. The method comprises
XX milling a dried sample to produce microparticles; suspending these in
XX a liquid buffer; extraction of nucleic acids from the microparticle;
XX passing nucleic acid-containing solution through a molecular sieve;
XX passing nucleic acid-enriched fractions through an anion exchange
XX chromatography material; and recovering fractions containing purified
XX nucleic acids. The nucleic acids are sources for sequences that encode
XX either operons involved in a metabolic pathway (specifically polyketide
XX synthetase) or polypeptides, particularly for production of therapeutic
XX or agricultural compounds, especially polyketide antibiotics. The present
XX sequence represents cosmid a26g1 (non-coding strand). The sense strand
XX encodes different polyketide synthases.
XX
XX Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;
XX
XX Query Match 52.5%; Score 47.8; DB 22; Length 42717;
XX Best Local Similarity 70.3%; Pred. No. 2, 1e-05;
XX Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
OY 1 CCGAGCGCGACGAGGATGGCGAGCCGCTGATCGGAGGCTCGGCATCAACC 60
DB 3525 CCGATGCGCGCGGTGACGCGCGATCGATTCGTGCGGTGATTCGCGAGTCGCGCTCAACC 3466
OY 61 ATGATGCGCGCGCTGACCGCGGTTGACCGCGCC 91
DB 3465 ACGGCGCGCGAGCAACGCGCTCAACGCGCC 3435
RESULT 11
AAF14651
ID AAF14651 standard; DNA: 77536 BP.
XX

```

AC	AAA14651;	
XX		
DT	08-Aug-2000	(first entry)
XX		
DE	Nucleotide sequence of the FK-520 biosynthetic gene cluster.	
XX		
KM	FK-520; polyketide synthase; PKS: gene cluster; immunosuppressant	
KM	Streptomyces hygroscopicus var. ascomyceticus; immunophilin;	
KM	FK-506 binding protein; polyketide compound; transplant rejection;	
KM	graft-versus-host disease; uveitis; alopecia universalis;	
KM	autoimmune chronic active hepatitis; inflammatory bowel disease;	
KM	multiple sclerosis: primary biliary cirrhosis; scleroderma;	
KM	neurite outgrowth; nerve regrowth; Parkinson's disease;	
KM	Alzheimer's disease; stroke; traumatic spinal cord; brain injury;	
KM	peripheral neuropathy; ss.	
XX		
OS	Streptomyces hygroscopicus.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	complement (412..1836)
FT		/*tag= a
FT		/note= "fkpM gene"
FT	CDS	complement (2020..3579)
FT		/*tag= b
FT		/note= "fkpV gene"
FT		3969..4496
FT	CDS	/*tag= c
FT		/note= "fkpR2 gene"
FT	CDS	complement (4595..5488)
FT		/*tag= d
FT		/note= "fkpR1 gene"
FT		5601..6818
FT	CDS	/*tag= e
FT		/note= "fkpE gene"
FT		6808..8052
FT	CDS	/*tag= f
FT		/note= "fkpF gene"
FT		8156..8824
FT	CDS	/*tag= g
FT		/note= "fkpG gene"
FT		complement (9122..9883)
FT	CDS	/*tag= h
FT		/note= "fkpH gene"
FT		complement (9894..10994)
FT	CDS	/*tag= i
FT		/note= "fkpI gene"
FT		complement (10987..11247)
FT	CDS	/*tag= j
FT		/note= "fkpJ gene"
FT		complement (11244..12092)
FT	CDS	/*tag= k
FT		/note= "fkpK gene"
FT		complement (12113..13150)
FT	CDS	/*tag= l
FT		/note= "fkpL gene"
FT		complement (13212..23988)
FT	CDS	/*tag= m
FT		/note= "fkpC gene"
FT		complement (13452..13662)
FT	misc_feature	/*tag= n
FT		/note= "ACP6"
FT		complement (13761..14394)
FT	misc_feature	/*tag= o
FT		/note= "KRC"
FT		complement (14517..15294)
FT	misc_feature	/*tag= p
FT		/note= "ER6"
FT		complement (15438..16587)
FT	misc_feature	/*tag= q
FT		/note= "dehydratase domain (DH) 6"
FT		complement (16587..17820)
FT	misc_feature	/*tag= r
FT		/note= "acylttransferase domain (AT) 6"

```
FT misc-feature complement (17820..19053)
FT /tag= s
FT /note= "KS6"
FT complement (19116..19326)
FT /tag= t
FT /note= "ACP5"
FT complement (19464..20097)
FT /tag= u
FT /note= "KR5"
FT complement (20241..21420)
FT /tag= v
FT /note= "DH5"
FT complement (21420..22653)
FT /tag= w
FT /note= "AT5"
FT complement (22653..23892)
FT /tag= x
FT /note= "KS5"
FT complement (23992..46573)
FT /tag= y
FT /note= "fkB gene"
FT complement (24163..24373)
FT /tag= z
FT /note= "ACP4"
FT complement (24997..26146)
FT /tag= aa
FT /note= "DH4 (inactive)"
FT complement (26146..27430)
FT /tag= ab
FT /note= "AT4"
FT complement (27430..28684)
FT /tag= ac
FT /note= "KS4"
FT complement (28750..28960)
FT /tag= ad
FT /note= "ACP3"
FT complement (29092..29740)
FT /tag= ae
FT /note= "KR3"
FT complement (29869..31018)
FT /tag= af
FT /note= "DH3 (inactive)"
FT complement (31018..32185)
FT /tag= ag
FT /note= "AT3"
FT complement (32185..33439)
FT /tag= ah
FT /note= "KS3"
FT complement (33505..33715)
FT /tag= ai
FT /note= "ACP2"
FT complement (33823..34480)
FT /tag= aj
FT /note= "KR2"
FT complement (34606..35749)
FT /tag= ak
FT complement (28750..28960)
FT /tag= al
FT /note= "DH2 (inactive)"
FT complement (35749..37144)
FT /tag= am
FT /note= "AT2"
FT complement (37145..38296)
FT /tag= an
FT /note= "KS2"
FT complement (38371..38581)
FT /tag= ao
FT /note= "ACP1"
FT complement (38677..39307)
FT /tag= ap
FT /note= "KR1"
FT complement (39442..40609)
FT /tag= aq

FT misc-feature /note= "DH1"
FT complement (40609..41842)
FT /tag= ar
FT /note= "AT1"
FT complement (41842..43093)
FT /tag= as
FT /note= "KS of extender module 1 (KS1)"
FT complement (43144..43660)
FT /tag= at
FT /note= "ACP of loading domain"
FT complement (43777..44629)
FT /tag= au
FT /note= "ER of loading domain"
FT complement (44974..46573)
FT /tag= av
FT /note= "CoA ligase of loading domain"
FT /tag= aw
FT /note= "fkB gene"
FT /tag= ax
FT /note= "fkB gene"
FT /tag= ay
FT /note= "fkB gene"
FT /tag= az
FT /note= "KS7"
FT /tag= ba
FT /note= "AT7"
FT /tag= bb
FT /note= "AT7"
FT /tag= bc
FT /note= "ER7"
FT /tag= bd
FT /note= "DH7"
FT /tag= be
FT /note= "KR7"
FT /tag= bf
FT /note= "KS8"
FT /tag= bg
FT /note= "AT8"
FT /tag= bh
FT /note= "DH8 (inactive)"

FT misc-feature complement (29869..31018)
FT /tag= ae
FT /note= "KR3"
FT complement (29869..31018)
FT /tag= af
FT /note= "DH3 (inactive)"
FT complement (31018..32185)
FT /tag= ag
FT /note= "AT3"
FT complement (32185..33439)
FT /tag= ah
FT /note= "KS3"
FT complement (33505..33715)
FT /tag= ai
FT /note= "ACP2"
FT complement (33823..34480)
FT /tag= aj
FT /note= "KR2"
FT complement (34606..35749)
FT /tag= ak
FT complement (28750..28960)
FT /tag= al
FT /note= "DH2 (inactive)"
FT complement (35749..37144)
FT /tag= am
FT /note= "AT2"
FT complement (37145..38296)
FT /tag= an
FT /note= "KS2"
FT complement (38371..38581)
FT /tag= ao
FT /note= "ACP1"
FT complement (38677..39307)
FT /tag= ap
FT /note= "KR1"
FT complement (39442..40609)
FT /tag= aq

FT misc-feature complement (40609..41842)
FT /tag= ar
FT /note= "AT1"
FT complement (41842..43093)
FT /tag= as
FT /note= "KS of extender module 1 (KS1)"
FT complement (43144..43660)
FT /tag= at
FT /note= "ACP of loading domain"
FT complement (43777..44629)
FT /tag= au
FT /note= "ER of loading domain"
FT complement (44974..46573)
FT /tag= av
FT /note= "CoA ligase of loading domain"
FT /tag= aw
FT /note= "fkB gene"
FT /tag= ax
FT /note= "fkB gene"
FT /tag= ay
FT /note= "fkB gene"
FT /tag= az
FT /note= "KS7"
FT /tag= ba
FT /note= "AT7"
FT /tag= bb
FT /note= "AT7"
FT /tag= bc
FT /note= "ER7"
FT /tag= bd
FT /note= "DH7"
FT /tag= be
FT /note= "KR7"
FT /tag= bf
FT /note= "KS8"
FT /tag= bg
FT /note= "AT8"
FT /tag= bh
FT /note= "DH8 (inactive)"

Query Match 52.5%; Score 47.8; DB 21; Length 77536;
Best Local Similarity 70.3%; Pred. No. 2.2e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 CCGACGCCGACGAGATGGCGACGATCTGGGCGCTGATCCGGGGCTGGGCATCAAC 60
DB 53053 CCGACGCCGAGCGCAACGCCGCGTGTGCGGTCCGGGGTTCGCCGCTCAAC 53112
QY 61 ATGATGCCCGGTCGACCGGCTGACCGGCC 91
DB 53113 AGGACGTCCTCAACGGGCTGTCGGCC 53143

RESULT 12
AAS08693 standard; DNA; 109519 BP.
XX AAS08693;
AC
XX
DT 26-SEP-2001 (first entry)
```



```

FT      /product= "Evrv"
FT      complement (42714..42717)
FT      /tag= az
FT      complement (42810..43799)
FT      /tag= ba
FT      /product= "Evrv"
FT      complement (43807..43811)
FT      /tag= bb
FT      complement (43799..44866)
FT      /tag= bc
FT      /product= "Evrx"
FT      complement (45014..45760)
FT      /tag= bd
FT      /product= "Evry"
FT      complement (45767..45770)
FT      /tag= be
FT      complement (45962..46714)
FT      /tag= bf
FT      /product= "Evrz"
FT      complement (45952..45956)
FT      /tag= bg
FT      complement (47156..49234)
FT      /tag= bh
FT      /product= "EvsA"
FT      51627..52715
FT      /tag= bi
FT      /product= "EvsB"
FT      51629..51622
FT      /tag= bj
FT      52889..53557
FT      /tag= bk
FT      /product= "EvsC"
FT      53554..54207
FT      /tag= bl
FT      /product= "EvsA"
FT      complement (54362..55117)
FT      /tag= bm
FT      /product= "EvsB"
FT      complement (55125..55128)
FT      /tag= bn
FT      complement (55135..56094)
FT      /tag= bo
FT      /product= "EvsC"
FT      complement (56100..56103)
FT      /tag= bp
FT      complement (56184..56813)
FT      /tag= bq
FT      /product= "EvsC2"
FT      56961..58709

```

```

Query Match      52.5%; Score 47.8; DB 22; Length 109519;
Best Local Similarity 70.3%; Pred. No. 2.3e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY      1 CCGACGCCAAGGATGGCGACGCATCTGGCGCTGATCCGGGGCTCGGCATCAACC 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      27466 CCGACGCCAAGGATGGCGACGCATCTGGCGCTGATCCGGGGCTCGGCATCAACC 27525
QY      61 ATGATGCCCGGTGACACGGGTTGACCGCGCC 91
      | | | | | | | | | | | | | | | | | |
DB      27526 AGGACGGCGCGACCAACGCGATCATGGCACCC 27556

```

```

RESULT 13
ABN76829
ID      ABN76829 standard; cDNA; 330 BP.
XX
AC      ABN76829;
XX
DT      08-JUL-2002 (first entry)
XX
DE      Human synthase-like ORF1776 cDNA, SEQ ID NO:3551.
XX

```

Human; ORF; open reading frame; ORF; drug screening; diagnosis; disease monitoring; cytokine; cell proliferation; cell differentiation; immune modulation; haematopoiesis regulation; tissue growth; angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic; thrombolytic; tumour inhibition; bodily characteristics; fertility; behaviour; cancer; proliferative disorder; neurological disorder; cardiovascular disease; immune system disorder; organ transplantation; tissue growth disorder; tissue regeneration disorder; diabetes mellitus; hypothyroidism; cholesterol ester storage disease; infection; vunerary; vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic; neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic; cardiac; hypotensive; antihypertoid; antiinflammatory; immunomodulator; dermatological; analgesic; vitucide; antibacterial; fungicide; gene; ss.

Homo sapiens.

WO200190366-A2.

29-NOV-2001.

24-MAY-2001; 2001WO-US17076.

24-MAY-2000; 2000US-206690P.

(CURA-) CURAGEN CORP.

Leach MD, Shinkets RA;

WPI: 2002-106200/14.

P-PSDB: ABP32803.

Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and disorders related to organ transplantation

Claim 1; Page 1148; 2508pp; English.

Sequences ABP31028-ABP35561 represent 4534 novel human proteins designated ORF (open reading frame) 1-4534, and sequences ABN75054-ABN79587 represent cDNAs encoding them. The invention also encompasses polypeptides at least 80% identical to the ORF1-ORF4534 (collectively referred to as ORFX) proteins, polynucleotides at least 85% identical to the ORFX nucleic acid sequences, vectors and host cells comprising ORFX polynucleotides, the recombinant production of ORFX proteins, antibodies specific for ORFX proteins, methods of detecting ORFX polynucleotides and polypeptides, methods of screening for modulators of ORFX expression or activity, and methods of screening individuals for a predisposition to an ORFX-associated disorder. The ORFX proteins of the invention have a wide range of biological activities, such as cytokine, cell proliferation, cell differentiation, immune modulation, haematopoiesis regulation, tissue growth, angiogenesis, activin or inhibin activity, chemotactic/chemokinetic activity, haemostatic activity, tumour inhibition activity, receptor/ligand, antiinflammatory activity, tumour inhibition activity, and antiinfective activity, and may also be involved in the determination of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester storage disease, and infectious diseases caused by viral, bacterial, fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 413.361 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065

Sequence: 1 ccgacgcgcgaacgagtgcc.....tcgacgcggttcacgcgcgc 91

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	100.0	58733	1	AF217189
2	91	100.0	68750	1	AF210843
3	91	100.0	68750	6	AR193029
4	91	100.0	68750	6	AR199551
5	91	100.0	68750	6	AR199559
6	91	100.0	68750	6	AR199567
7	91	100.0	68750	6	AR201097
8	91	100.0	68750	6	AR208671
9	91	100.0	71989	6	AR172664
10	65.4	71.9	16124	1	AX024384
11	65.4	71.9	16124	6	AX024277
12	55.8	61.3	49736	1	AF319998
13	54.2	59.6	104326	1	AB070940
14	52.2	57.4	10910	1	AX024383
15	52.2	57.4	10910	6	AX024276
16	51	56.0	22838	1	AF081920
17	51	56.0	27541	6	AX211706
18	51	56.0	123580	1	AF263912
19	51	56.0	125401	6	AX211739
20	49.4	54.3	30000	6	AX350262
21	48.4	54.3	66808	1	SA0421825
22	48.4	53.2	27522	1	AB070942
23	47.8	52.5	671	6	AX153711
24	47.8	52.5	764	6	AX153716
25	47.8	52.5	3792	6	AX195972
26	47.8	52.5	4306	6	AX153796
27	47.8	52.5	6462	6	AX153794
28	47.8	52.5	34071	6	AX153790
29	47.8	52.5	39314	1	SGR300302
30	47.8	52.5	42717	6	AX153789
31	47.8	52.5	77534	1	AF235504
32	47.8	52.5	84985	1	SNR278573
33	47.8	52.5	109519	6	AX195929
34	47.8	52.5	113193	1	AF357202
35	46.4	51.0	330	6	AX310566
36	46.2	50.8	758	6	AX153712
37	46.2	50.8	5292	6	AX153793
38	46.2	50.8	24045	1	AB070941
39	46.2	50.8	30000	6	AX250261
40	46.2	50.8	30000	6	AX250263
41	46.2	50.8	32870	1	AE007122
42	46.2	50.8	42603	1	AF188287
43	46.2	50.8	65140	6	AX211705
44	46.2	50.8	107379	1	SHCPIR
45	46.2	50.8	107379	1	SHCPIR

ALIGNMENTS

RESULT 1
AF217189
LOCUS
DEFINITION
AF217189
Sorangium cellulosum putative transposase gene, partial cds;
putative transposase gene, complete cds; epothilone biosynthesis
gene cluster, complete sequence; putative membrane protein gene,
complete cds.
ACCESSION
AF217189
VERSION
AF217189.1 GI:745354
KEYWORDS
POLYANGIUM CELLULOSUM.
SOURCE
POLYANGIUM CELLULOSUM
ORGANISM
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 58733)

AUTHORS Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B.

TITLE Cloning and heterologous expression of the epothillone gene cluster

JOURNAL Science 287 (5453), 640-642 (2000)

MEDLINE 20115953

PUBMED 10649995

REFERENCE 2 (bases 1 to 58733)

AUTHORS Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C.

TITLE Isolation and characterization of the epothillone biosynthetic gene cluster from *Sorangium cellulosum*

JOURNAL Gene 249 (1-2), 153-160 (2000)

MEDLINE 20293058

PUBMED 10831849

REFERENCE 3 (bases 1 to 58733)

AUTHORS Julien, B.

TITLE Direct Submission

JOURNAL Submitted (16-DEC-1999) Kosmos Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES

source

1. 58733

/organism="Polyangium cellulosum"

/strain="SMP44"

/db_xref="taxon:56"

<1.992

/note="orfA"

/codon_start=3

/transl_table=11

/product="putative transposase"

/protein_id="AA62378.1"

/db_xref="GI:7453555"

/translation="VRGH/ENFADFGVPRVLYDNKNAVERHGDALRPHPTLAL
SADYREPRVAVARNEKGRVEGARIVREGEFEARAYADLDGLNQAWEVSSAL
DRSVEEDRATVROAFDD3RSVLIRHPDTPPDHERVEVGKTPYARPLNYSVAL
DRPRTLVLAIDISOVRIADGNOIVATHVRSMDGQOIEPHEIQRORLDEKRRRER
GLDILARAASSQAFRLIVAEKGNVGSATIRLQILDVGAELERATLVEVLEDTI
HICAVRQVIDRRSRERHLPPVSIPTVREGHALLVTPHSLTTITDALKKPTP"
989.1501

/note="orfB"

/codon_start=1

/transl_table=11

/product="putative transposase"

/protein_id="AA62879.1"

/db_xref="GI:7453556"

/translation="MTDLPTETKRLKSLGELGLACWELQADKPKLREVLAIERE
RHKSLERLRKNSVAFKPMTDSSMPKRIDREAVDDLYDSRYADLLEEVYTRRD
AOKPLSLTKKAAVMQGVPRHACVTVLVRILVHRAVEIEAESYRLKEAKELNAT
RTKORRTKKH"

1998.6263

/gene="epoA"

1998.6263

/note="type I polyketide synthase; epothillone PKS loading module"

/codon_start=1

/transl_table=11

/product="EpoA"

/protein_id="AA62880.1"

/db_xref="GI:7453557"

/translation="MADEPIERAEDPIATVAGACRLPGGVIDLSCPTLLRSDRV
GOYAPERMDAAAFDDPDIDAPKGTPTVRAPIIDVACFEDSAFSGICSRRLRSDPRV
LLLEVCEALEENAIAP:ALVGTETGYTIGTSEYEAALPRATASKEIDAHGGLGPM
PSVAGRISTVIGLRGPCVAVDTRASSILVAHLACQSLRSGECSTLACGVSILMSP
STLVKSTKRALITDGRKAFSAADGFGEGCAVYVLRSGARADGDRITLAVIG
SAINHGSAGLVVPMNG:SOEIVLKRALADAGCAASGVYEAHGCTTGDPLEIQA
LNAVIGGRDVAITLIGSVKTNLGHREVASGITGLIKVLSLOHGOIIPAHLHAQALN
PRISMGDRLTITVARTIYWDNTPPRAAGVSFGMSGTNAHYVLEERPAATCTPPAE
RRABDLVLSKRTAAADP:HAARLDHLETPSOQLGVAFSLATTRAMEHRLAVAT
SSSEGLRALDAAGQGTTPPGVVRGIADSSAGKAFLETTGGAGQTLGKGLYVMPAF
REAFEDLCVRLNQLDRP:LRVVMMAEPAVDALLDQTAFTQPALFTFEVALALMSP
MGVEPELVASHIGELVYACVAGVSELEDAVPIVAARGRLMOALPAGCAVSIAPPA
DVAAPAAVPAASVIAVNGPDQVIVLAGAGQVPIIAAAMAARGARTKALHVAHAPS
PLAAPHLEAGGRAVESY:YRRPSITVLSNLSGKAGIDEVSSPGYWRHAEVYRFADG
VKALHAAGAGTFVEGPK:STLGLIVPAQLDAPRALASSRAGRDEPATVLEALGGLM

AVGLVSWAGLFPSSGGRVPLPTYPWQREXYWIDTKAADARDDRAPGAGHDEVEKG
GAVGGRDRARLDHPPESGRRKEVAAADRPRLIEDPVLIDRLEKTERNAPG
LGEVEIVDAAGLSFNDVOLALGMKVPDDEKGNPDLILGECAGRIVAGSEVNLV
VGOPVIALSAGAFATHTTSAVLVLPPOALSTBAAMVAVITAMVADIGIARLP
GERLIIHAATGCGVGLAQAQVAGVNAHTACTPERKATLSEIGRYVYDSRDFP
ADVAKWTGGGVAVLNSLSELIDSFILNRSHGRVEIGKDCQADNOLGKPLFR
NLSEFSLDLRGMLEPARVRALEIELLGLAGVTPPPATVLPARAAVADAEAFRLTO
AOHKGKVLTLGDEVOIR:PTHAGAGPSTGSDLDLRLASAPARAALAEAFRLTO
VSGVLRTEIKVGALEALFRIGDLSMAVELRRIRASLAKIKSTFTLSPTIALIT
ONLIDALATLSIERVAENLRAGVDSFVSSGADQDWELIAL"

6260.10492

/gene="epoB"

6260.10492

/note="non ribosomal peptide synthetase; epothillone PKS module 1"

/codon_start=1

/transl_table=11

/product="EpoB"

/protein_id="AA62881.1"

/db_xref="GI:7453558"

/translation="MTINOLNELHOGVLAADGRLOTQAPKALNPMLARISEH
KSTILMLRORLPAESIVPAPARRHVPFPLTDIOGSYVIGRTGAFVPSGIAHREYD
CTDLDVARLSRAFRKYVARDM:RAHTLPPMAGVIEPKYDADIEITDLGSDRTEA
RLVSLDRAMSHRITDTERPPLVHVAVRDEDOOTRLVSLIDLVLSLISLIFKDWL
SFYEDPETS:LPVLESLRSDYVLALESRKSEAHQRMDYWKRRVAVLPPPM:PKMD
PSTLREIRFERHTEQMLPDSMSRLKORVGERGLTPGVILLARSEVIGRMSASPRFTL
NITLFNRLPVHPVNDITGDETSMLVLDITDTDKSEORAKRIGOLMEAMDHCVS
GTFORPARVILGICORGALFPVYLTSALNOOVGVN:SIORIGPVYTSOTPOLLDH
QIYEHGDDLVLANDYDGVFPPLLDMLAEYATLRLTEEPMSQMKCSIPPAOLE
ARASANTNLSLEHTLHGLFAARVQDLPMQALVARSATLYTELSRRSRKATLR
EOARPNITLVAVMEGMEQVAVLVLESGAAVYIDDLPER:HYVLDHGEVYL
LTPQDLKXLSMPG:IORLRLVPSAGVEGDDQ:PPMPIDTSPSLAVIYITSGSTGYPK
GWNIDHGAANTLIDINEREPIGPGVVALSLSLSDVYDFGLAAGTIVVDA
SKLRDPAHMAALIEREKVYMWNSVPLVHSHSGRSDLSASRLSLSGDWTIPV
GJGGEQATIRGVYSISGATASITSGYPRANDP:SMASITGYRPLR:NOTFHYLD
EALPRVWPGQYITGGVLA:GVWDEKTKNSFLVHEETGRLYKTDGGRYLPD
GNIEFMREDNOQIKLRGYRVEIGEIEETLKSHPNDVAVIVPGANDANKLLAAVVP
EGTRRRAAEQDASIKTERVDARAHAAKADSLSGEBOVQFLRAGHGRDLGKVPDL
TGLVPRAGGLDVYARRSRVTFLEAP:LPVEFEFRFLSCSSVPPDGAALPKFXYPSAG
STYPOVTVAAKSGRIEGVDEGFYYHHPFHRLLKXDSHIEGAVHPONFVDEBAA
FGLIFVGRIDATISLIGSLRSECLIEACGMQALMEQAPSCVIGCPYQGPDEBYR
PVLDLRHSDVTVHGM:AGRVDPFOFVCTIGODSSRRKTRTAPRAGROHREILRD
FLTRKLPETVPTVFEVLDLPLTSNGKVDKRLRERKOTSSRHSGHTRAPDALEEI
LVAVGEVSDCLELEVQVQSDVFDGATS:HVNRKSLRORLREIRAITELFOYPNIGS
LACGLRDSKDLDEORPNMODRVARRRKGRRS"

10639.16137

/gene="epoC"

10639.16137

/note="type I polyketide synthase; epothillone PKS module 2"

/codon_start=1

/transl_table=11

/product="EpoC"

/protein_id="AA62882.1"

/db_xref="GI:7453559"

/translation="MEBOSSAIAVIGMSGRPGARDLDEFMRNLBDGEAVOFSEGO
ELAEASVDPALVLDPNVVRAGSVLEVDNRDAFPICISREALIMPOHRIECCAME
ALNACVDPAAVAGS:GYVAGAMSSYVLSNHEHRAAMRWOMPOTLIGNKCYLAT
HVTIRLNRKPSITSVOTACSTSLVAHAHLKSLDDECMHLAGGLTIV:IPHAGVY
ABEGIT:SPDHCAFPDAAKANGCTIMNGCCVYILKPLDRALSDDPVRAVILTAALRVF
GARKIFETPAGGQAALMEALATMGCAAS:TOYLETGCTLLGDA:ETALARVE
DRASIRPACISVSTIGCHESSAGIAGLITVYAL:LEHROLPSLINFESPMSIDF
AASBPVFNSTLKMVNGSPRRAGVSSFGIGYNAHVVE:EEAPAAKTLPPAARSAEL
FVYSASAAALDAARLRDHDLOHGLSLGVAESLATTTRSMERLMAAPSRAL
RGLDRAAQRQTPGA:VNRKGSFNPVAVFVPGGSCWVGKROL:LEEFTHAL
SACDRA:IOEAGMSLAEIAELAGSSQOIERIDVQVLELA:AAFAFLRSMVGGVDV
VIGHSMEGAAVAAVAGALSLIEDVAV:ITCRSRSLRRI:SCGEGAAVTEL:SLAEAEALR
GVDYRVA:VANSNPSRSTVLSGEPAA:IGVLSLNAAGV:CCRAKV:VAVASPOVDLR
EDIL:LAAGLRPAAPAAVPMRSYVTGMAVAGPELAGVMMNNT:ROP:PRF:AVYQAOLOG
GHGLPYEMSHPLITTSVEBMRAARACAGVSLRGGDERPAMLEALGALIMAOCTP
VPWGRLEPPAGGRVPLPTYPWQREXYTEAPKASAGDRRGRAGGPHLLGEMQTLST

```

QSTRLEMTTLDLKRPLWGDHVRGAVVFPAGAYLEMAISSGABALDGPLOITDYY
LAELAFAGDAVAVIYQVYVTEPQSGRLQFOIASRSGHSEFVHAGALLRVTE
VPAGITLSAVARLQASMPAAATVATLTEMLOVYBPAGQALREHGGALLGRRLP
DAAGSAEYRLHRLPALLDPCFOYVGSILFAGGCEATWYVEVGSILDRPSCELMCHA
RVNHGRQTPDROGADFNWVDSGAVVAEVSGLVAQRLPGGAREEDDWLELEMER
AAVGAATAKNAGRMWLLGGGGLGALRSMLEGGVAHVAAESNNSAAGVALLAKAF
DGOAPTAVVHLGSLDGGGELDPGLGACALAPRASVDSPALDELGRVIAHEADLCARDYD
OLAGMGFRDAPRLMLTRGAQVAGDVSTQAPLGLGRVIAHEADLCARDYD
PTRPGEIGALLIELLADDAEVALRGCECVATVRROETPRRGIESCVPTDY
IRASTYIVYTGGLGGLSVAGMLAERAGHILVYGRSGASVEORAAVALEANGAR
VTAADVADRQERILREVTTSGMPLRGVHAAGILLDDGLMOQTAARFKVAPK
VOGALHLHALTREPALSFVLYASGVGLSGGQGNAAANTFLDALHHRRAQALPA
LSVDWGLEAEVMAAODRGARLVRSGMRLTDPDEGLSALARLDELHHRRAQALPA
PRLMWEIYPAASRMLSRVTAHRASGAPGADDLRLAARPEARSGLLEPLLR
AOISOVLETPGKILEVDPLTSLGMSIMGLERIRITAMGLGITVPAILLMTYPIVAA
LSGHLARACCAAPESPHTTADSAVETLEBSODDITOLIAKFKALT"
16134..37907
/gene="epod"
/gene="epod"
16134..37907
/feature="type I polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="WTRGPTAQNPLKQAIIIOLEERLGLAQALETERPIAY
GIGCFPGGADAPFAFELLDARDAVPLDRMALVGAPEVPAHAGLITEPIDC
EDAFEFISPREARSLDPQRLLEVAEMGEDAGIIPRSIDSGTGVFAFTADYD
RTVALPREEDAYSATGNMLSTIAGRLTGLQGPCTIVDTACSSILVAHLACRS
LRAGESDLAGVSTLLSPDMMEAAARTOALSPGRCTDASANGFVGGGGLVY
LKRLSDAQRDRIAMALIRGSAINHDSGTCTANVTAQCEVTALEARSHEVAGV
DYERHGTSLSDPIEYEAIRATYGPAPRSQGTRECVLCAKNTICHELAAAGALIK
AALSLTHERITPRNLNFTLRPIRLGSSALALATEPVPPTDRIRFAGVSSFEQSIT
NAHVLTLEAPAVEMLPAPERSAELVLSGSEGLDQAARLRHLMHSELIGDY
AFSLATRESAMTHRLAVSTRREGILALSAVACOTPAGARCIASSRGELILFT
GOGAOTPGMGRGLCAAMPAREAPREDCTLDRELDPLREVMNAEAGSARELILDT
AFTOPALFAVEYALTALRMSWGPBELLVSGELVACVAGVSELDGQVIAVARG
RLMGISLGGAMVSLGAPAEVAAVAPHAAMVSLAAYNGPEQVITACVBOAVQALIA
GFAAGVTRKRLHVAHFHSPIMEPELEEFGRVAASVTVRRPSEVLNLSKVVTTDE
Query Match 100.0%; Score 91; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGAGCGCACGCGGATGGCGACCGCTGATCCGGGGCTCGGCCATCAACC 60
|||||
DB 16975 CCGACGCGCACGCGGATGGCGACCGCTGATCCGGGGCTCGGCCATCAACC 17034
|||||
QY 61 ATGATGGCGCGGTGACCGGGTTCAGCGCGCC 91
|||||
DB 17035 ATGATGGCGCGGTGACCGGGTTCAGCGCGCC 17065
|||||

RESULT 2
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS Polyangium cellulosum.
SOURCE Polyangium cellulosum.
ORGANISM Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangiineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnanow,M.,
Gottschalk,J., Mayo,J.M., Engel,N., Toupet,C., Straumann,A., Cyr,D.D.,
Gottschalk,J., Mayo,J.M., Hu,A., Gott,S., Schmidt,J. and Ligon,J.M.
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from Sorangium cellulosum So ce90
JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source
1..68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"
1..1826
/feature="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDLAPRATSGARVLDGGRATERVLAERGIEDLRALREH
LRIOGGSFHCMLGDLTVBELADQPLASISPHARSLRHPMTSDAMLVDGPALV
RWLAARGAPRLREYEEERERARTAOEARRLMLAAPCPADLPREDANGLPLGP
MSPEVAERLRASATPELACALLAMLTGTAGPMGYPAYEMLPENLLIGGLPT
ALAAASRGTSSEALRGARLPASWEVSSKSGOLNPEALMERLARTIRAMGNADN
LSRPERARAIARERLRQAPAPRAGAGLAVAGYSSSGRLSGVTDGALYSGDGN
IVMPQGRISPVVLAGTDPPELAPPLPSOMLFAHNAAGTISKVTEGSPLYARN
QARPSLHVHAGFAMVNOAVNPDERCAPVVOYSTIMEHEHPTPCLHPPASAFS
LACDESHLWMLCSGRLELRRHHRGAPRSPAYLGEHITAAWTYPSLNTHTVL
WADPRVVALVQMDKRTGVEPIVLETRHPPAVVSEDDIPALDGOFPSRDMVHEIR
SGASVTVADYORLMDRDPVLMNRGLFEFTINDHLILARS"
complement(1900..3171)
/feature="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to DQ-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAAPGALGAEEREGPOGDAGALAPAREMAEVA
AGKMGDAVLAARGDNDVDAVYTELGGSPMRDRTFRTASMTKATATAVMVLE
EGKLDIDSPVRLMPELARKVLRIRGSDIDETVPAERPTVRDMLPTMGFISFDA
SSPIORATDELGLVNAQVPMTPHCPDEWIRLGLPLMHQPGQAMWYNTGSLVQYL
VGRADQGFDAFVRKRLIAPJMKRDTHVADKLARAGCGITIDEQTEKTRDRL
GAESAAYASPAPFPGAGLVSTVDYLLFAPMLNGVYHEGRRLISAASVEMTADHL
TPAQKAASSFPFGFEFTGMCGYMAVTPADPAVSEVPGYMGDGFSTWINDGREL
IGIVTOSAGFELSGALERFRSVYVATESA"
3415..5556
/feature="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF24240"
/db_xref="GI:6724240"
/translation="MHGTEROVLSTVLTALITVTAAGELARLRPQVIGELRG
GVAGPSVAGALPQHRALFOEPRAVGVYLSIGTALILLIAGIEVDGILRKEA
RPGALSALGAIAPLAAGAASALVLDPLPSGLFLGLIVLSTVAVTAKVLIREESM
RRSYNOVTIAGVSEVAANVLVAMTSSYASASPALAAARSALLASGLPMVIVGR
LTHLMRWADATRSGOVSLVILVTFLLAALYORGLHLPLDAFALGVLVNLAPRT
NRLPLDGVOTVAGLFAFVFIAGMRVDVSLORPAMGVALLIARATKAVPAA
LCARLIGIRGSEALVANGIMKGGTDLIVATVEYGLISNEAYTAAVAVATVTA
SPALLIWEKRAPPTQESARLREERARRAYIPQEVILVPIYAHALPGATDIVES
IVASKRKLGETVIDITELSEVQAPSPASAGEASGLARLGLRVGIWRORELGS
IOALIRASRDLDLVITGARSFARAGMFGRLQDAIVORASNVLVVVGDPAPARAS
ARLILVPIIGLEYSAAADLAHVALADADLVLLSSQTPGAVVMDRSPARAV
ARVYDEAFREGRRLGVSSRVHVGAPSPDEITRELARADYDLLVAGCYDHGLGRLL
VLGSVESVVRVPVALLIYAHGSTRQYR"
complement(5612..5992)
/feature="ORF4"
/codon_start=1
/transl_table=11

```

/product="unknown"
 /protein_id="AAF26907.1"
 /db_xref="gi:6724241"
 /translation="MDKPI3TRCAIAEGV1PGSGNCEPEQMOTSHETACILNADRDRA
 QVAIVYVSDRBDPAGPRVIVPARTRIVRPNNDLPEEPIIPRODYASVIESDAPIVV
 OHTRLDSRQAEALLSTIAYTDRE"
 6226..6675
 /note="ORF5"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAF26908.1"
 /db_xref="gi:6724242"
 /translation="MKHYCTGRRCGRIRGHTLGLLASMALACGGSGSEKTYOGRILAP
 GADAVYTDVDDPATTTIAYDVVHLSPERLEKSEKFEVWQNPSPSPRRVGVLD
 INADSRCKLKEVYIPVAKPELLITAEKQSSPQSSAAVIGTISVg"
 7610..11875
 /gene="epoa"
 7610..11875
 /gene="epoa"
 /note="EPoS A"
 /codon_start=1
 /transl_table=11
 /product="polyketide synthase"
 /protein_id="AAF26919.1"
 /db_xref="gi:6724253"
 /translation="MADRITERAAEDPIAIVGASCRLEPGVYIDLSGFWTLLEGSRDYV
 GRVPEEDDAAMADDPDIAPGKTPVITASFLSOVACFDASFFGI3PREALRMDPAIR
 LILEWCMBLENAIAIAPSLAVGTETGVFTIGPSEYEALALPOATSAEIDAHGGLGTM
 PSVSGRISYVALGLRGPICVADTAASSSLVYHLACQSLRGSESTALADGVSILGSP
 STVLWLSSTYALARDGRCIKAFSAEADGREGGAVVYIKRLSGRADSGRIILVING
 SAINIDGASCLITVPNGSJOETIYKRALADGCASSVGYEANGCTTIGDPTETIQA
 LNAVYGLDRDVAITPLIGISVKTNLGHPETASGITGLKLVVLSLDHGOI1PAHLAQAOLN
 PMISGDIARLIVTRATPPMDPMTNPRAGVSFGMSGTNAHVYLEAPATCTPPAPE
 PRAEILVLSAFTASALDAQJSLRDLHLETYPSCQIGVAFSLATTRSGAMEHRLVAVAT
 SREGIRALDADAAGQST3GAVRSIADSSRCKLAFLEFGGAQOTLGMGRGIDYWSAF
 REAPFLCYRLTNOELDRP..REVMNAEPASVDAAL1DDQTAFOPLFTPEVALIAIAMS
 MCVPEELVAGHSIGELVAICYAGVFSLEDAYFLYAAGRILMQALPACGAMSTIEPAPA
 DVAAVVAHAAVSITAAVNAAPDQVVIAGAGPVHAI1AAMARCARTRKALVSHAFIS
 PLKAMLEAFAGFEVEYGSP3TLLGLVPLVACMPAPRALLASAGDEPAPVLEALGGLIM
 AVAGVSNAGLEPFGSGRRVPLPTYPWQOBERYWDITKADDAARGDRAPAGHDEVEEG
 GAVRGDRSRAIRLHPHRESGRREKVEAGRGPRLEIDEGVDHILVYTERBARG
 LGVEVINADAGLSFNQY2IALGMVPPDLLPEKPNP1LLIGSECGRIVYVAGEGVNCIV
 VGOPIVIA1SAGAFATHVITSAALVLPKPOALSAIEAAMPVAYITVAYALDIRLARLP
 GERVI1HAATGGVGLAAV2MAOHVGAETHAATGPERKRAYLESIGVRYVSDSRDRPV
 ADVNRMTGCEGVVDVNLISGELLIDKSNLILSRSGREVEIKRDCYADNDLGLPEFLR
 NLSFELVLRCKMMLERPARVRALLLEELIGLCAAGVFPMPPIATPIIARVADAFRSMQ
 AOHGKILVLTGDEPVOYRIPTIHAGAGAGSTGDRDLRLIASAAPARAALAEFTFRO
 VSOYLRTPEITVGAELFTRLGMDSLNAVELRNRIEASKLSTTLELSTSPNIALLA
 QNLDAIATVLSLEKVAENLRAGVQNDVFSSGQDWEIITAL"
 7640..8920
 /gene="epoa"
 /note="Region: beta-ketoacyl synthase/malonyl-CoA
 decarboxylase"
 9236..10201
 /gene="epoa"
 /note="Region: acetate acyltransferase"
 10529..11428
 /gene="epoa"
 /note="Region: enoyl reductase"
 11549..11764
 /gene="epoa"
 /note="Region: acyl carrier protein"
 11872..16104
 /gene="epob"
 11872..16104
 /gene="epob"
 /note="EPoS P: contains condensation/heterocyclization,
 cysteine adenylation, putative Mbd(P)H oxidase, and
 peptidyl carrier domains"
 /codon_start=1
 /transl_table=11

Query Match	100.0%	Score 91:	DB 1:	Length 68750:
Best Local Similarity	100.0%	Pred. No. 1.5e-12:		
Matches 91: Conservative	0:	Mismatches 0:	Indels 0:	Gaps 0:
OY 1	CCGAGCGCCGAACGGGATGGCGACCGCCATCTGGGCGCTCGGCGCATCAACC	60		
DB 22587	CCGAGCGCCGAACGGGATGGCGACCGCCATCTGGGCGCTCGGCGCATCAACC	22646		
OY 61	ATGATGGCCGGGTGACCGGGGTGACCGCGCC	91		
DB 22647	ATGATGGCCGGGTGACCGGGGTGACCGCGCC	22677		
RESULT 3				
LOCUS	AR193029	68750 bp	DNA	linear
DEFINITION	Sequence 1 from patent US 6346404.			
ACCESSION	AR193029			
VERSION	AR193029.1	GI:20238994		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 68750)			
TITLE	Schnupp, T., Ligon, J., Madison, G., Molnar, I., Zirkle, R., Cyr, D., Dawn, and			
JOURNAL	Gorlach, U.			
FEATURES	Genes for the biosynthesis of epothilones			
SOURCE	Patent: US 6346404-A 1 12-FEB-2002;			
	Location/Qualifiers			
	1..68750			
	/organism="unknown"			
BASE COUNT	9596 a 22456 c 25359 g 11159 t			
ORIGIN				

Query Match 100.0%; Score 91; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 60
|||||
Db 22587 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 22646
|||||

Qy 61 ATGATGCCGGTTCGACCGGTTGACCGCGCC 91
|||||
Db 22647 ATGATGCCGGTTCGACCGGTTGACCGCGCC 22677
|||||

RESULT 4
LOCUS ARI199551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI199551
VERSION ARI199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNALS Patent: US 6355457-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 91; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 60
|||||
Db 22587 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 22646
|||||

Qy 61 ATGATGCCGGTTCGACCGGTTGACCGCGCC 91
|||||
Db 22647 ATGATGCCGGTTCGACCGGTTGACCGCGCC 22677
|||||

RESULT 5
LOCUS ARI199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION ARI199559
VERSION ARI199559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNALS Patent: US 6355458-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 91; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 60
|||||
Db 22587 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 22646
|||||

Qy 61 ATGATGCCGGTTCGACCGGTTGACCGCGCC 91
|||||
Db 22647 ATGATGCCGGTTCGACCGGTTGACCGCGCC 22677
|||||

Db 22587 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 22646
|||||

Qy 61 ATGATGCCGGTTCGACCGGTTGACCGCGCC 91
|||||
Db 22647 ATGATGCCGGTTCGACCGGTTGACCGCGCC 22677
|||||

RESULT 6
LOCUS ARI199567 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI199567
VERSION ARI199567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNALS Patent: US 6355459-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 91; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 60
|||||
Db 22587 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 22646
|||||

Qy 61 ATGATGCCGGTTCGACCGGTTGACCGCGCC 91
|||||
Db 22647 ATGATGCCGGTTCGACCGGTTGACCGCGCC 22677
|||||

RESULT 7
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNALS Patent: US 6358719-A 1 19-MAR-2002;
FEATURES Location/Qualifiers
source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 91; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 60
|||||
Db 22587 CCGACGCCAACGGGATGCGACCGCATCTGGGCGCTGCGGCGCATCAACC 22646
|||||

Qy 61 ATGATGCCGGTTCGACCGGTTGACCGCGCC 91
|||||
Db 22647 ATGATGCCGGTTCGACCGGTTGACCGCGCC 22677
|||||

Db 22647 ATGATGCGCGTGACCGGGTTGACCGCGCC 22677

RESULT 8
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoehlones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 91; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 60
|||||
Db 22587 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 22646
|||||

QY 61 ATGATGCGCGTGACCGGGTTGACCGCGCC 91
|||||
Db 22647 ATGATGCGCGTGACCGGGTTGACCGCGCC 22677
|||||

RESULT 9
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epoehlones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source Location/Qualifiers
1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 100.0%; Score 91; DB 6; Length 71989;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 60
|||||
Db 16975 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 17034
|||||

QY 61 ATGATGCGCGTGACCGGGTTGACCGCGCC 91
|||||
Db 17035 ATGATGCGCGTGACCGGGTTGACCGCGCC 17065
|||||

RESULT 10
LOCUS AX024384 16124 bp DNA linear BCT 15-SEP-2000
AX024384

DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024384
VERSION AX024384.1 GI:10184588
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
REFERENCE 1 (bases 1 to 16124)
AUTHORS Myxococcales; Sorangineae; delta subdivision; Myxobacteria;
JOURNAL Beyer,S. and Mueller,R.J.
Patent: DE 19846493-A 13-APR-2000;
FEATURES
source Location/Qualifiers
1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN

Query Match 71.9%; Score 65.4; DB 1; Length 16124;
Best Local Similarity 82.4%; Pred. No. 2.5e-06;
Matches 75; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 60
|||||
Db 472 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 531
|||||

QY 61 ATGATGCGCGTGACCGGGTTGACCGCGCC 91
|||||
Db 532 AGGATGCGCGTGACAGGGTTGATGACCC 562
|||||

RESULT 11
LOCUS AX024277 16124 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
REFERENCE 1 (bases 1 to 16124)
AUTHORS Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
JOURNAL Beyer,S. and Mueller,R.J.
Patent: DE 19846493-A 82 13-APR-2000;
FEATURES
source Location/Qualifiers
1..16124
/organism="Polyangium cellulosum"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN

Query Match 71.9%; Score 65.4; DB 6; Length 16124;
Best Local Similarity 82.4%; Pred. No. 2.5e-06;
Matches 75; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 60
|||||
Db 472 CCGACCGCGAAGCGGATGCGGACCGCATCTGATCCGGGCTCGGCCATCAACC 531
|||||

QY 61 ATGATGCGCGTGACCGGGTTGACCGCGCC 91
|||||
Db 532 AGGATGCGCGTGACAGGGTTGATGACCC 562
|||||

RESULT 12
LOCUS AF319998/c 49736 bp DNA linear BCT 28-MAY-2001
DEFINITION Stigmatella aurantiaca myxaliamd biosynthetic gene cluster,
ACCESSION AF319998
AF319998

VERSION AF319998.1 GI:14210834
KEYWORDS
SOURCE
ORGANISM Stigmatella aurantiaca.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Cystobacteriineae; Cystobacteraceae; Stigmatella.
REFERENCE
AUTHORS 1 (bases 1 to 49736)
Slakowski, B., Nordisiek, G., Kunze, B., Blocker, H. and Muller, R.
TITLE Novel features in a combined polyketide synthase/non-ribosomal
peptide synthetase: the myxalimid biosynthetic gene cluster of the
myxobacterium Stigmatella aurantiaca S9a15
JOURNAL Chem. Biol. 8 (1), 59-69 (2001)
MEDLINE 11182319
PUBMED 11182319
REFERENCE 2 (bases 1 to 49736)
Slakowski, B., Nordisiek, G., Blocker, H. and Mueller, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2000) Mx. GBF, Mascheroderweg 1, Braunschweig
38124, Germany
FEATURES
source
1..49736
Location/Qualifiers
/organism="Stigmatella aurantiaca"
/db_xref="taxon:41"
complement(2..1306)
/note="ORF1"
/codon_start=1
/transl_table=1
/product="putative aspartokinase"
/protein_id="AAK57182.1"
/db_xref="GI:14210835"
/translation="MRPIGSAQAGRGTSALALIVKXGTSVGDTERIKNVAKCIA
AOKAGHDVVVVSAMSGETNRLKIVSITERNPREDVDVVAAGEVSGIVLALIO
TOKKRAVSFLGHVIRITDTSFNRKARISIDARIVDALKKKNIVVVGCGODEHGN
VTTRGRGSSOTTAALAAALAKADACEITTDGYTTDPNNVPAPAKERLAIEMLE
LASTGAKVLQTRVSEFAKIVPLWAKSVFDDPGTLLCEEDAMEDVLSGIVADKN
EAKIANGVPPVGAAKIFGALDEOSIVDLIVQNVSKDRDTVTFTVAGADLTAK
DVAKRIASGVAEVEDDQSVKSIYGVGNRNSGVAKMFTVLACGAVNVQMISTS
EIKVSCVHSNVTETLAVRSLHTAFGLDKPEAGAVSEVYSLKG"
1363..2895
/note="ORF2"
/codon_start=1
/transl_table=1
/product="putative histidine ammonium lyase"
/protein_id="AAK57183.1"
/db_xref="GI:14210836"
/translation="MSRPRMLIDGDTLKEITLLOVAKHTVTELAIPAAAYKAAARDL
VDVAAAGDTSYGTINTGFTLAVERIDKIDRELRNLILSHAAGVGPLPAPARVL
LLRCNVLAKGYSQIRPETLALALEMNRDVPVPERGVSAGSDLAPLHNLVFI
GEGEAFYKGERLPAQAALERAGLPVYLKEGALVNGTOAMCAVGTLLQRAEMLA
DLADLAGMTLEGLGSHKPFIPETIODVRAHEGOKAACAHRELIDASAIYESHNC
KVDPYSILRCMPQVHGAAAREGLSFARILVEINSAIDNPILVPEYTERIVSGGNHGO
PVSLADYAAAMALVQLSISRERYBOLVNPALSGIPPLAKNSGLNSGFTLAQVTSAA
LVAESRVLSPASVDSIPSSAGREDHVSMTGALDKGVADPFRCCIAITELVAQA
LDYRQPTIRAGKGPQAAAYELIRSKIPJMKEDRELHRDIAAVSALIDSGELNAVTRAT
GQ"
complement(3275..7822)
/gene="mxaa"
complement(3275..7822)
/gene="mxaa"
/note="non-ribosomal peptide synthetase; contains a
condensing, adenylation, and PCP reductase domains"
/codon_start=1
/transl_table=1
/product="mxaa"
/protein_id="AAK57184.1"
/db_xref="GI:14210837"
/translation="MSYRELITTLAAGVRLRLBEGDKLOVEASKAVYLPELAEISAN
KADILALRKHSREETGSELGOLAPREVPILSSGGERLWFLDRLSPTGSOYVNHG
LARGALDTKALRSLDELVRHRLVTHPEVSGSPROVIVSPDTGCVALTIVLPGI
TESOREEIVRSDELAAOOPDILAKGRLRTVYALGPPDPAITYTKHHTITDKSLG
VFVELSMLYASFVKGQAAALPPVSMQPADSLERAMLAAGESGARLISAATWKGKLG
PPLQDPVDAVSTSHRGATVPLALSPALSDALRELAIRGCSFFVYVLFAPASLLH
RSGQVDFGVGTVIANKGSVPADLIGFIANTLALRCDLISGPTFSOMLARAHHVYLEA

gene
CDS
1..7819
complement(7819..11520)
/gene="mxab1"
complement(7819..11520)
/note="PKS: ketosynthase; acyltransferase"
/codon_start=1
/transl_table=1
/product="mxab1"
/protein_id="AAK57185.1"
/db_xref="GI:14210838"
/translation="MRAGAGGHPLTGVSFSTSVOKENFMESALSAREPYLADHRG
GAVVVAATYEMLSARRELSGEKAYLNDTSREGLILPEGGERYVOMLSSEGG
ALSRQLSQKASTERNAPACMAHYVGOIRIEGSPSTQTPSEBIAIERSEVYAOA
QHYAEIASRGVAAGPAGFQGVQOIMKGRREALGHRLEPPLAARSASVIRIPHALDAC
OLVAAAYQEDLTPDPAQPAVVALAHLEHSPGTEVCHRLRPSBARGVYEFDL
VLRDASGRILVEALGILRYOQDLPAEAGHGDLEFALEMRASAVQVAPAGPASGR
MLIADSGELADYESOLREHGEVIRVDLAGAGGRRVVDATSPQADVLGGVGE
GSARGVHTLGLPREVPDITMCGGCVHLYOVALASVTSAPLMLITRVHHNGSG
RSADYVQASLWGLGRTLAHESSEKCTRVDSVANSMTGDAKMLRLELVADDEBEI
SLRDGRHVGRIVKSGPGRAPSEAVVPAGGRRPFLIERAQGRLEPRARARPPEGEL
LIEVEVALGPGRGEGTGPERRGGAEGTGCGRVVAIGEGRNVAAGEERIALVASGL
GSHVSPAGCTAIRPSSSSHEAALAAITPANYALTEHLGRLEKRGVWLGATSGI
GRAAIKLARAGAEVFAVATDEORASIRELSAGAOVMPDSVSPSYRLDGGVGE
LAIYAPGAVLELERSLPTLSBGRVYDLSASGASARQVDSANAKVYVQDPRARRPER
FASIMQEKAAFPENAGGLPSAADAALKALPKPKARVVPADARRLRADGTYITVGL
GGLTAAVAKMNVEDGAGHLVLVGRDVTLPBQOEYVTALEAGARVRAADVSDRAO
IARILSEIABEGAPLRGVITHAGVLDVLDLQOYERRRVMAKRVILGNNLHVLTRE
APLDFEYLYSAAAMFCAPOGCGVYAGNAFLDALAHHRALGLGLSTINMGPFSEVGL
AAAOANREERLAOAGSDSLTPAEGVTILGRLLDGVYQMAVPLELRKRVPEYPRAS
SPMLSEIYPAAGGQGEGBARVALLLETIRTAQPEAKALRQVREOLARVLRIDSNR
IDPERPLOSRTGLDMLGELRNRLYSGVGLPASLITKHPTLALCALCHLIGVMDRT
LAETLQAAGAGPVDSDADDEKEVFL"
complement(11487..14303)
/gene="mxab2"
complement(11487..14303)
/note="PKS: dehydratase; enoylreductase; ketoreductase;
ACP"
/codon_start=1
/transl_table=1
/product="mxab2"
/protein_id="AAK57186.1"
/db_xref="GI:14210839"
/translation="MSKOEKPAEQPSALQORALLVEKQARLDAVERAREPIAVIGL
GCRPPGAVDSSSYWKILRDVDALREVPESSRMVPGHFDSTROPVAKMGVTRGGLFD
DVHEHDAEFFGTSIPREALASLDPOORLVLEPESELENNAGLSTPDLASGTFMGVWS
SDYMARILKENDATREFDGYATAGYSVPPRVSIVYGLQGPCAPVPDTCSSSYVAGG
LASELSRAGESNLLAGGVNLLSPETMIGCSMOALASDRCRPTPASADGVYAGG
GCYVYKRLSNAQDGSILALINGSANHHGASGGLTVPRNGPQOAVYQADLNARI
APALVGTLEAGGTPIPLGDTLELRLAGLVAGKRPADPFFTVGSKNIGLIEPAAVSI
AGVITKTLISLQHKETIPHLHFRTPNPHEMDRIPARVPEKVPMPAPKGRGRIIACVSS
GLGSIHNVYILEEAPASTAPAAVEEDKAOVLVYSAFTPAKSLIAEWSFLIENEG
GSLADISYTSALSRPHEHRLALAAKSKQELIEQLEAFRAKSHAGLSGRRGTCIGQ

KIVFVPGGSGOMLNGORLYAENVTFREAIIDMCHESIASTRYDWSLRDLITNPGQOPR
 NMDIIVOTPTLALOVALLAWMLGLEPHIAVAGSMGAAVAHAGALSLDPAARI
 CERSKILTRPFGSGAMAYVDI.SREQAEVIRIACAEVIRIAGVNSPSTYSLDGPALK
 EYLERLEREVYCRVYKUDV/SHSROMDOLRODILDRMAIRITPGSAVLYSTVGCAD
 IUGRELSAPIYRNLRQPYLADAVORLQDSNFIETISPIPIVLPFYDAMIAREDT
 SGRGLVAPSLINDOEWPSSJSSVGLFTFVIAWGTSGYIPTSVAGLRPLRTGSSAG
 TGMSCOOLRRRSGCALAREVIPS"
 complement(14300..29878)
 /gene="macx"
 /gene="macx"
 complement(14300..29878)
 /gene="macx"
 /note="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
 ACP regions"
 /acp_regions"
 /codon_start=1
 /transl_table=11
 /product="Macx"
 /protein_id="AAK57187.1"
 /db_xref="GI:14210840"

gene
 CDS

/translation="MANRDESNELSPAKRMVLAEKQTRLNVEGAKEPIAYGM
 ACRFPOADNVEAFMRIDGVDVAREIPEIRPKDEVADIGGAMWGFLDSVDFD
 PGFEGISPREAVRMDPOORTLLEAVNEALDGLDVKLSGRSGVFICAGNDYHGM
 OYREPTGDFATGATVAAAVLSGRLSYLPILOGPSLYVDTACSSLSVSLHACOSLRN
 RECNALAGVNLISPOSTLYLSKIOALSPPGRSKAFAPASANGPTRGCCITVILKR
 LSDLADGDHIIATIRGSAI NODGKSTGLTNNVLSQALIRQALESSGLKADQSYV
 EAHGCTPLGDIIEAEALRHTYGVPRPDGSGVIGKTNVGHLESAGIAGIMKVL
 AMRHOSIPPHLHLKQVNPRIQLESGALITPTRLTPWKASGPRAGVSSFGISVIAH
 VIEEAPAVAKDALQONRPIVPLPSARSREALVALAERYAERLGRGSGSLDVCYA
 SAHRTHTHRLVADSRREITRERLAFRAGPOSGAMSGRAKAFRRITVYPPGQ
 AORACMGRAIMSEPEFRDLYKVKDALOPHGMSTLDEIADDEORSRLAEDVONPV
 LRAVEVALTALMRKCIETPVAVGHSMGEVAAHAGALSLEDAVILITHRKMLRT
 GGRGATMAAVELSEAEVIRIPWNGKIMLGLINGRSQVLSGEPAVEAVEALDKA
 GVTRVWKMDVSHPTVEIISRELEGRITARTIRADIPICSTVGELEAVEALDAA
 HMGNNLRNFRPFAAAOHLASDHCTFEVSEPHVPLPVEQOTMSAGNTEGVILSLR
 RHEDEHRTLSLGLSYSHVYANMRVYRGOLVPLPTVPMOREFMEPAVIGVR
 AARTSGHPLTGERIDLSAGGEVSRLOMIGPKHPLIAEHRGCVPLPAAAMDM
 ALSAKQITGEPEVSLSDYSALLSPPEBEVYLOMLSTERGASAFLECAPKRE
 TPSDELTMALHAGSLRHG CASVEKAAVADDELTKARCGEESVBAHQOTKEIGLE
 YGSPFRVARELYRANGELARVOLBELTAEBERYSLHAPLALDGCQYVLAALLEGDS
 EGVYVVALSLRYLVRPEALMLSHVRIVSAGEAGSNTIECDVTLNQGGRVAVEVR
 RLVCKKIAASAPADEVRMITSAGHEPQOORARATGSSRTAGRLVADNRAGIOS
 LASGLAERAGCVVVEAGPSYERLEADRYRINPSKLEDFQALLKEAFSOPSCKGVH
 LYGLDVASVNTPRESEFERALSSTSVYLLDANMORGRDTPRLMLYTTQVYIASG
 EOPGSPAAPLVGLAVITHEHRELCAHVDLSAOPSPREIASLVEELDGEDEDIA
 LRGAARVHVARLGRGISGRAAEPREFTAGTYLITGGLGLEIAVAKMPTIGKARRIA
 LTGRGAPSEEAQKVLGELFANGAEVYRADVTSADVARLGLSIDSELPARGVHA
 AGVINDGVITQDEPERLRVLAKEVQANNLHVQGRPLELFVMPSSAAMSGSPGQ
 GNYAANAMFNVLAHHRBAGLPGLSVMNPMWADGLAVATNRGDRLLRGVGNPP
 AQAELALGRLLVQEAQVIVPDLRQMRFPYLSAQSPLSVLMQEKYSTAPSRGS
 ARELLAAEPEDRKGLTECYLRBOGRITLMEBSRIEPPQPFEDLGWDSLUTIELNRK
 IESGKLTLTSLATLTYAVPVQALTYTYLYGKLEPVEAEARPTVPRPTQTEPRVSVGAP

Query Match 61.3%; Score 55.8; DB 1; Length 49736;
 Best local Similarity 75.8%; Pred. No. 0.004;
 Matches 69; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 CGGACGGCAACGGATGCGACCGCATCGGCGCTGATCGGGGCTCGGCATCAACC 60
 Db 29046 CCGATGGCGTGGGACGGGATGCATCATCATCTCGCGCATCGGCATCAACC 28987
 QY 61 ATGATGGCGGTGCGACGGGTGACCGCGTC 91
 Db 28986 AGGATGGCAAGTCGACAGGGCTGACCAAGC 28956

RESULT 13
 AB070940 104326 bp DNA linear BCT 22-JAN-2002
 LOCUS Streptomyces avermilitilis o1gomycin biosynthetic gene cluster.
 ACCESSION AB070940
 VERSION AB070940.1 GI:15823967
 KEYWORDS
 SOURCE Streptomyces avermilitilis DNA.
 ORGANISM Streptomyces avermilitilis

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 JOURNAL
 1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
 Shinoue, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osone, T.,
 Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
 Genome sequence of an industrial microorganism Streptomyces
 avermilitilis: deducing the ability of producing secondary
 metabolites
 Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
 2 (bases 1 to 104326)
 21477403
 Ikeda, H.
 Direct Submission
 Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
 Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1
 Shirokane, Minato-ku, Tokyo 108-8641, Japan
 (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
 Fax: +81-3-3444-6197)

FEATURES

source

gene

CDS

/organism="Streptomyces avermilitilis"
 /db_xref="taxon:33903"
 complement(1..1134)
 /gene="mcma"
 complement(1..1134)
 /gene="mcma"
 /note="methylnonyl-CoA mutase"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB69185.1"
 /db_xref="GI:15823968"

/translation="MTGYSTADANAREQLVSGTRAVPVFDLPTRAGLDSAPAA
 RLQVGRGVAVSIDDMRYLFGGLRDVPSLAADAPAPLLVYOLVAEHBGAPQ
 RLQAVRGVFEKPEFLARGVPELRSILANVYACAEIPRMOCSVCGHLAGT
 GADPLAEVAVAGGIEVYRAVATGIDQVRYAPRLTLCVIGTGTGRENRAKLRATR
 VMTRRANERFGSGRRAPRITYTPTRDGLATLACTEDIEGALVADRIRLGGITLYL
 LKRAAGQGLRPHGRDPAVRPHPHQDLSLQAKOTERLAKLRAMRVOPAVDEALR
 complement(1131..1886)
 /gene="ppta"
 complement(1131..1886)
 /gene="ppta"
 complement(1131..1886)
 /codon_start=1
 /transl_table=11
 /product="phosphopantetheinyl transferase"
 /protein_id="BAB69186.1"
 /db_xref="GI:15823969"

/translation="KRAHVEFADVPVSSIHPREALYODAVESRRREFATGHCARRA
 LASLGVPEGLATPRGAKRMPAGVCGATITHCAGTAAVAVARAPGIRAVGIDAEPRP
 LPEGVLEAVALPAEQVYKERTARKHGVWGLFSAKESVYKAMPATGRPLGEAA
 LIDLDPVSTFEFRVRLPTLRAAEPLELDLHGRWLVRHGIIVTAVVADGPGHDTFHP
 SPYDRTKHSAPGRSADPRGCRPRTRATDRPYA"
 complement(1910..2875)
 /gene="simx4 homolog"
 complement(1910..2875)
 /gene="simx4 homolog"
 complement(1910..2875)
 /codon_start=1
 /transl_table=11
 /product="Simx4 homolog"
 /protein_id="BAB69187.1"
 /db_xref="GI:15823970"

/translation="MSQDRHDSSTAALGSPSHGRLLAIVSHVRNDRITVODLHP
 GCAEDMLIVAGDVASVADITWALSILDRFAKVIWAPGNHELMAPGPDRLRGERK
 YRLVLDLGRALGVTPEDPYVWNGAGVAVAPFLILDYSEFLPPGASKSHALRLA
 DDAGVCTDEPFLHDPYRSREARMCARARATRLDALPDLPLVLYVHFPVLEPPT
 RYLRHPEPALNCGITRTADMHRIYRAAAVVRHRLHLPRIIRCDXVPHHEVSLGYREV
 QRYGRPDPVPIITPAPAPVAVPVVAPPPHRTASPADRRP"
 3213..3971
 /gene="olmc"
 3213..3971
 /gene="olmc"
 /codon_start=1

gene
CDS
/transl_table-11
/product="thioesterase"
/protein_id="BAB69188.1"
/db_xref="GI:15823971"
/translation="MSANRYDGNIMVROFHPAPTAATKLVCLPRAGSASVFLPVSNA
LAPGVDTALIOYPGKODRREKCIDNLAEDAVTALEPADPPLVFGHSMATLA
EVALRLERKIVPLALPASGRAPSRHRESVHLRGDDEIATLKTLSGRSVLTGD
EELIMVLPALRSDYKAAETRYAGARLAPPIAHFVGEDPRVSLDEAQAAMDHTTG
HPELRSHGPHFYLNDOAPVIEIARSSVSA"
4643..7438
/gene="olmri"
4643..7438
/gene="olmri"
/codon_start=1
/transl_table-11
/product="luxr-family transcriptional regulatory protein"
/protein_id="BAB69189.1"
/db_xref="GI:15823972"
/translation="MLIGRQEDIRLRAAVSVSRGSRRAVLVESAGCGKSEILED
LVRRVESGLIALATAIAERNTPLSVLRSLVNGAPFPDILVRRPHSDAAARMTS
TEGRDKARKDAAMYELMRTECEVRLAQRAPLVVAVDDIHGDADSLPOLHLAHNC
RNPVLLTFTOSLAGSGAPRTFGAEILROBNPHIRLPPRCRDIADMLAEQGRADG
OGTVERFYAVSGNPLLRLLLEDALDEPDYDADVGPFRGQATGRAVLACLDGG
EKAHEVAAGLAVGGAFTHELLRLDLSAALTENLRALRAAGLVGGSFQHPAKA
LIADERRAVSYLELAHAASADPRGRTEIKIRLGSYTRVSADARRHVDALRGFA
GILTPSDTGLAALLAHGRLEQARSVLEQGRPGDGDADRTAINGCTEGLLETHID
ATPDLVLAHVMTLLAGVERALHMGDTFLREKRRDAGRAVIAVAEABALSGN
LGRBROGILATPQLAMVODGAPRIKISLRLOGLASPEARLGLINRAVELSEALR
RIELSTLTDLADHRETRQTAQAAVAVRAROLAKSCGAPRRIRISGPDAAAGSE
LREGEPRNAWDGHLSESERRYAGLAASGYTRNEISRLYITMSVVEQLTPYVRKTK
INREELIAEAAGAY"
7658..10486
/gene="olmri"
7658..10486
/gene="olmri"
/codon_start=1
/transl_table-11
/product="luxr-family transcriptional regulatory protein"
/protein_id="BAB69190.1"
/db_xref="GI:15823973"
/translation="MLASAGCGEGVIVNMGVGTALLDAFLRKATSGARVCA
AGSSAETAEERGVQLPPTGPGAGRLAIGSRDAGLADRLFDALCOELR
RPTVLAVDADQLADASLRLRYTRRLTAPVMVVLTEPTGPALPSPQAEILRP
OCRRLGLPLSRDGYTQVSAQVSDRYRLADPHALISGNPLLVHALLADHRAHR
LSPVSPVTVSVSSASASASASASASASASASASASASAKALAMAYRDEPVL
DVACGFAVIGESAPPARVACLVDGADADVMTADTAGLGGTRLSALVGALE
HVDVTRAEIRRAARILHGDCAPIYVARHLIAPAEPAAGRTLLAALEKAOEGR
EELIDICRLAGQATTEQERAAVMAVKIGMIDIRLVSPMIDELVALRDLDS
GDAMITYHLAMODEPPEAVESLTLAAAGDAGEPVAELHFEVRLMRSCSELEVS
VPGSNPTPTPTPLRVRSPASVAGEVYGLTGPDDHIVAMAEELILRCRSEYV
EALPAALALAVYAGPRPAFWCDGLLOAGTRSTGSAVILISGIRAEIALROGALPA
VEQYARSALSAISLPGWGVAVGSPLLALVHSTDSGSGVAGAMLNADVDMETRH
GLTYHARGHHLAVDRPAALIDDELTCGGLAKWGMADVFTLPWRTSAQADALGN
PGOARALVREQLDGRGSGSTARARVALLRLAATGEPERRALLRESVNLSCGDVE
LITGEGSSASLTAAERVTALAAIGHNSNRIGKRLFTTKSTVQHILTRYVRKLGVRNR
ADLDGLAGTNIGTGPAMAMASAVAG"
10933..11271
/codon_start=1
/transl_table-11
/product="p450-like protein"
/protein_id="BAB69191.1"
/db_xref="GI:15823974"
/translation="MEVGGQITRAGGVLTCTSTANRDEGAFMDPELIDLRDASHL
SFSGIHOCIGOSILARAEPQIALETLLRLPLGLRLAASPEELRFRKGVGYKELPD
PCPALPPDRP"
11458..29898
/gene="olmri"
11458..29898
CDS

Query Match 59.6%; Score 54.2; DB 1; Length 104326;
Best Local Similarity 74.7%; Pred. No. 0.00085;
Matches 68; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 1 CCGACGGCGCAAGGATGCGACCGCATGCGCGTCAATCCGGCGGCGCATCAACC 60
Db 57234 CCGACGGCGCGCCGGGAGCGGCACCCCGTCTCGCGGCGTCCGGCGGTGAACC 57293
QY 61 ATGATGGCCGCTGACCGCGGTGACCGCGCC 91
Db 57294 AGGACGGCGGCTGACCGCGCC 57324
RESULT 14
AX024383
LOCUS AX024383 10910 bp DNA linear BCT 15-SEP-2000
DEFINITION Sequence 81 from Patent DE19846493.
ACCESSION AX024383
VERSION AX024383.1 GI:10184587
KEYWORDS
SOURCE
ORGANISM
Polyangium cellulosum.
Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 10910)

AUTHORS
Beyer, S. and Mueller, R.J.
JOURNAL
Patent: DE 19846493-A 13-APR-2000;
STATEMENT OF WORKING COPY (SOWC)

FEATURES	Location/Qualifiers
source	1. .10910

BASE COUNT	1852 a	3124 c	3613 g	2321 t
ORIGIN				

Query Match	57.48;	Score 52.2;	DB 1;	Length 10910;
Best Local Similarity	74.28;	Pred. No. 0.0038;		
Matches 66;	Conservative 0;	Mismatches 23;	Indels 0;	Gaps 0;

QY	3 GACCGCGCAAGGGATGCGCAGCCGATCTGGTGCCTGATCCGGGGCTGGGCCATCAACCAT 62
Db	3000 GACGCGCAGCGCGCACCGCGACCCTTCCTGTGCGTGCTCAGAGAGCACGGCGATCAACACAC 3059

```

QY      63  GATGCCCGGTCACCGGGTTGACCGGCC  91
          ||||| | ||| | | | |||
Db      3060 GATGCCCGGTCACCGGGCTACCGTCC  3088

```

LOCUS	AX024276	10910 bp	DNA	PAT 15-SEP-2000
DEFINITION	Sequence 81 from Patent DE19846453.		linear	

```

/organism="Polyang::um cellulosum"
/db_xref="taxon:56"
BASE COUNT      1852 a      3124 c      3613 g      2321 t

```

Query Match	57.48;	Score 52.2;	DB 6;	Length 10910;
Best Local Similarity	74.28;	Pred. No. 0.0038;		
Matches 66;	Conservative 0;	Mismatches 23;	Indels 0;	Gaps 0

Oy 3 GACCGGCAACGGGATGGCGACCGCATCTG3GGCGTGAATCCGGGGGCTCGGGCATCAACCATT 62
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3000 GAGCGCGACGGCGACCGCGACCCCATTCCT3GGGGTGTCAGAGAGCACGGCGATCAACCAAC 3059

```

Oy      63 GATCGCCGGTGCACCGGGTTGACCGCGCC 91
          ||||| | |||| | | | |||
Db      3060 GATGCCCGAGCAGCGGGCTACCGTGCC 3088

```

Search completed: November 5, 2002, 12:05:23
Job time : 561.361 secs